AD_____

MIPR NUMBER: 95MM5519

TITLE: Feasibility and Design of a Tri-Service Database
Architecture Allowing Service-Specific and Tri-Service
Reporting of Hospitalization Rates for Military Women's
Health Research

PRINCIPAL INVESTIGATOR: LTC John G. Meyer

CONTRACTING ORGANIZATION: Armstrong Laboratory/PS
Brooks AFB, TX 78235-5241

REPORT DATE: October 1995

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for public release;
distribution unlimited

The views, opinions and/or findings contained in this report are those
of the author(s) and should not be construed as an official Department
of the Army position, policy or decision unless so designated by other
documentation.

19960226 102

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	October 1995	Final (21 Nov 94 – 30 Sep 95)

4. TITLE AND SUBTITLE
Feasibility and Design of a Tri-Service Database Architecture Allowing Service-Specific and Tri-Service Reporting of Hospitalization Rates for Military Women's Health Research

5. FUNDING NUMBERS
95MM5519

6. AUTHOR(S)

LTC John G. Meyer

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Armstrong Laboratory/PS
Brooks AFB, TX 78235-5241

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, MD 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The overall goal of the project was to determine feasibility for a collaborative data communications infrastructure which would enable medical researchers from any service to access multiservice data. However, each service would maintain its own database. To these ends, the Air Force has examined the feasibility of (1) constructing a research-oriented comprehensive longitudinal database of demographic and medical data, (2) constructing a relational data model which might be usable by DoD researchers, and (3) constructing a user-friendly front end which would be able to access several databases at once and present them to the user in an integrated fashion. Internal cross mapping tables would bridge differences in variable naming and coding conventions. As a feasibility study, Air Force and Navy epidemiologists and statisticians extracted data from their own service data systems on hospitalization rates for women for eight diseases. The two services cooperatively agreed on common analytical methods and naming conventions and were able to produce a comparative report. The findings of this study are reported here. The feasibility study pointed out limitations of this approach, and the project team explored a variety of technical options which would more readily permit sharing of semantically-equivalent data among the services. A major accomplishment of this project was the creation of the Defense Medical and Epidemiological standard data set, which identifies the sharable epidemiological data currently available from the three services. The Air Force continues to pursue this goal in the development of the system initiated during this project. A user interface is being developed which allows the investigator to query more than one database for related information.

14. SUBJECT TERMS
Defense Women's Health, Database, Epidemiological Studies

15. NUMBER OF PAGES
64

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the US Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

___ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

___ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

___ In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

___ In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

___ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

PI - Signature   29 JAN 96
                              Date

Abstract

The overall goal of the project was to determine feasibility for a collaborative data communications infrastructure which would enable medical researchers from any service to access multiservice data. However, each service would maintain its own database. To these ends, the Air Force has examined the feasibility of (1) constructing a research-oriented comprehensive longitudinal database of demographic and medical data, (2) constructing a relational data model which might be usable by DoD researchers, and (3) constructing a user-friendly front end which would be able to access several databases at once and present them to the user in an integrated fashion. Internal cross-mapping tables would bridge differences in variable naming and coding conventions. In order to prove that it is feasible to present reliable joint studies, Air Force and Navy epidemiologists and statisticians extracted data from their own service data systems on hospitalization rates for women for eight diseases. The two services cooperatively agreed on common analytical methods and naming conventions and were able to produce a comparative report. Part of this final report summarizes the findings and presents the joint epidemiological study as an example of the type of research that could be accomplished with current systems and methods. In recognition of the limitations which these methods placed on producing such comparative studies, the project team explored a variety of technical options which would more readily permit sharing of semantically-equivalent data between the services. Key representatives of the Army, Navy, and Air Force met to address functional and technical issues concerning the identification and sharing of such data. A major accomplishment of this project was the creation of the Defense Medical and Epidemiological standard data set, which identifies the sharable epidemiological data currently available from the three services. This project also demonstrated that a common data communications infrastructure was not only feasible, but highly desirable to support future research. The Air Force continues to pursue this goal in the development of the system initiated during this project. A user interface is being developed which allows the investigator to query more than one database for related information.

Introduction

In recent years both the percentage of women in the Navy, Air Force and Army as well as the involvement of women in active deployment and combat has grown. In August 1994, representatives of the Navy, Air Force and Army medical research community met to discuss the feasibility of a Tri-service study of rates of hospitalizations, disease, and injury in servicewomen. Because of differences in format, naming conventions, and coding, there has been an inability to communicate across services about medical data for research purposes. A project to rectify this lack was undertaken and a voluntary effort to communicate medical data across services was *internally motivated rather than externally mandated*.

The overall goal of the project was to determine feasibility for a Tri-service collaborative data communications infrastructure to support the goals of the DWHRP. The project started by using existing systems, service-specific data sources, operating systems, and programs to the greatest possible extent to produce a pilot epidemiological survey of diseases prevalent in women. This involved extensive cross-mapping of variables and coding between the services and any existing DoD standard. However, the group identified a need to use more technologically advanced data extraction and analysis tools. Data would be kept in powerful relational databases amenable to data extraction with structured query language (SQL). Cross-mapping tables would bridge multiple databases. Joint studies would then become feasible without each service giving up its autonomy and service-specific conventions. Further, a front end would be developed which would allow a medical researcher to, without any systems knowledge, formulate a query which could then be automatically processed; the user could be served rapidly and, often, interactively.

During the initial meeting, it became clear that extant capabilities and strengths of each service differed. The Navy has had in place a longitudinal research database system containing comprehensive information on hospitalizations, demographics and career history for enlisted personnel since 1965, but its format is not amenable to SQL queries. The Army was in the process of developing a computerized operation-oriented medical surveillance system that included inpatient, demographic, and career data. The Air Force had an inpatient tracking system and was beginning the development of a medical research data system. The Navy and Air Force agreed to develop specifications for a Tri-service collaborative data communications infrastructure but recognized the reality of where each branch was in development. Consequently these two services agreed that they would jointly pursue a DWHRP effort to allow for these two systems to evolve to a place where they could create a common data structure that could be routinely used for both service specific and multiservice issues. The Army elected to concentrate its efforts on developing the Army medical surveillance database.

Over the next 12 months the Navy and the Air Force jointly pursued the efforts described in each of their reports.

The Navy concentrated on expanding and documenting their already extensive longitudinal database. They (1) expanded their database to include Marine Corps and Navy Officer Data and (2) extensively documented both the database, itself and the system of COBOL programs which parse, add, and retrieve data.

The Air Force initiated several tasks: These were (1) cross-mapping of variables between services; (2) developing a relational data model and ODBC compliant database design which could be used by both services; (3) building a graphical user interface prototype which was user friendly and could extract and combine data from both databases.

The two services produced a joint epidemiological study from cross-mapped existing sources to demonstrate proof of concept (Appendix).

Body

Methods and Materials

Systems Development

This report assumes that the reader has a rudimentary knowledge of databases and their structure. Even so, however, individual terms may have different meanings depending upon the reader's discipline. Further, there may be several terms which apply to the same object or concept. For example, a research thinks in terms of independent and dependent variables. A variable is, essentially, a symbol which represents the contents of a column in a data table. That column would be called a field or a data element by individuals involved in database design. In order to prevent confusion, we have constructed a 'Database Primer' which is presented in **Appendix A.**

User Requirements

This project addresses the collection, refinement, storage, and accessibility of information needed to accommodate scientists who deal with epidemiological and public health problems, issues, and forecasting. In order to determine what that information might be and how that group of researchers might best be served, individuals in OPHSA who were involved in such studies were brought together in focus groups to discuss their needs, list some of the most often used and critical variables used in their studies, and evaluate the contents of the Navy database. The disciplines represented were preventive medicine, flight medicine, occupational medicine, public health, behavioral psychology, sports medicine, and epidemiology. Researchers currently involved in studies of women's health issues were included in the user groups.

Similar input was solicited from other organizational entities including OASD/HA, and members of the Annual Tri-Service Epi-Board.

However, a more organized and thorough approach is required. Consequently, a questionnaire is being used to obtain more detailed information from potential users both internal and external to our organization. The questionnaire has been designed to address several issues. These are (1) what variables; *i.e.* kinds of information, the user might require currently and in the foreseeable future; (2) what platforms (DOS, Mac, Windows, UNIX, *etc.*) the individual used both at work and at home; (3) usage of computer resources, *i.e.,* software; *etc.* (4) problems the user had in obtaining data; (5) impact of being able (or not) to obtain the data;(6) degree of need and (7) to what ends might the individual apply the data if he/she could obtain it.

The recipients of the survey were chosen to include a broad cross section of Air Force medical organizations and research disciplines. Initially, the surveys are being distributed via U.S. postal service and electronic mail to a small subset of individuals in order to obtain enough information to refine the survey. Individuals are being asked to respond in a like fashion. However, this survey has been designed for the input of textual descriptive answers which can only be analyzed subjectively. Information derived from the initial round will be used to design a more objective survey. Eventually the survey will be accessible via the World Wide Web (WWW). These data will be placed in a Microsoft Access Database for easy retrieval and analysis.

System

An IBM model R-24 running an AIX operating system (IBM flavor of UNIX) is being used to accommodate major database structures. Sybase 11 is being used as the vehicle to construct prototype relational databases and as an information server. Currently, two prototype databases are envisioned with approximately the same relational structure. One database will house the Air Force Central Research Data Base (CRDB). The second will contain a subset of the Navy data which will then be in a format amenable to SQL queries.

Data

Obtaining Data

Although the ultimate goal is to encompass data for as many years as are available, for the purposes of the initial feasibility study, we chose to use a six year segment ranging from 1989-1995. This period encompassed a major deployment of women; *i.e.* Desert Storm and a post-deployment period. An extract of the Navy longitudinal database currently maintained as COBOL flat files was stored in a flat file format on IBM 3480 cartridge tapes Standard UNIX tools, including AWK, sed, cut and parse were used to parse the data into tables importable into Sybase. This is facilitated by the already well-defined event structure of the flat files.

Since the Air Force currently has no longitudinal database which combines demographic, career, and medical data, one must be constructed from raw data. After visits to the Air Force Manpower Personnel Center at Randolph Air Force Base and the Defense Manpower Data Center (DMDC), we determined that demographic and career data could best be obtained from DMDC. This source offered a number of advantages. The requested personnel variables were available in one file, and the accompanying documentation provided cross service mapping to a number of DoD standard codes. This documentation would facilitate our task of mapping the Air Force variables and code tables to the corresponding Navy data. Appropriate permission and documentation were exchanged. DMDC data is on IBM 3480 cartridge tapes. Each tape holds approximately 250,000 records. A total of 46 tapes were used to capture the DMDC data in flat file format. Two cartridges are required for every fiscal quarter of data snapshots

Inpatient data is available from the Standard Inpatient Data Record (SIDR) maintained in the Air Force Inpatient Data System at the Pentagon. It was obtained by OPHSA statisticians via a download from the SIDR mainframe as ASCII files. A subset of the SIDR data was extracted and is stored on hard disk on the MEDIC RS6000 in flat files.

Because both DMDC and SIDR data are quarterly snapshots of raw data, an event structure must be imposed upon the data. This will be done by establishing an initial record for an individual and then using changes in successive snapshot data for that individual to compose events. DMDC and SIDR data will be aligned via SSN, Name, and date of birth.

Mapping Data

Because each service has its own formats, codes, and conventions for addressing even very similar data, Navy and Air Force data cannot be used in parallel unless the variables from both services can be equated. Further, it is desirable from the standpoint of the individual service that such conventions not be abruptly rejected. Consequently, one of the team who has a long-standing familiarity with the data from both services equated the variables between services and also with the DoD standard codes. Data equality would be implemented by the use of mapping tables rather than actual transformation of the data.

Standardization of Data

Since this mapping is for proof of feasibility and for a limited number of variables, formal standardization has not been attempted. However, this project is committed to the use of DoD Standard Data Elements IAW DoD 8320.1-M-1, when they are available to help ensure commonality of form (syntax) and meaning (semantics) of data compared between and among the services. It is recognized that if the prototype system were evolved into an operational environment, the full procedures for data element standardization would need to be followed. The preliminary epidemiological study performed during this project illustrates many of the concerns which must be taken into account regarding the use of non-standardized data. This study served to point out many of the areas which need to be addressed in developing a true common data infrastructure. The DMED Data Standardization Workshop held during the course of this project made significant gains in this area. Another positive consequence of this meeting was the consensus by the services to use common data sources whenever possible, and to take advantage of cross-service conversion tables and interservice codes developed by DMDC. The Navy agreed to examine DMDC and SIDR data to construct their "retrofit database". This would ultimately simplify much of the data mapping and standardization with Air Force and Army data; these services already use DMDC and SIDR data

Translating Data into Relational Tables

Because the data files are so large, a subset of the Air Force data will be used to build a model database on the PC in Paradox. A database design tool (SmartER, TM,Knowledge Base Systems, Inc.) may be used to facilitate table design. This small database will also be used to test the user interface. Eventually, Sybase tables will be constructed and the six year subset of data will be distributed among them via the mechanisms developed for the small pilot.

Construction of Databases

Once the tables have been constructed, relationships will be built with included Sybase Tools. Because speed of retrieval depends upon which variable is used to sort the data (index), indexing will be based upon results of the user survey.

Communications

The Air Force Database will reside on an IBM-R24 in Sybase 11. Although direct access to the database might be had via Sybase Open Client, such access poses a considerable security risk. Consequently, any access to the database, itself, will be via a proxy server. Communication with the proxy will be via FTP messages using the Internet as medium of interchange.

User Interface

Platforms

The user interface is currently being developed to accommodate users running 16 bit Windows 3.1 running over DOS. However, not all users prefer this platform. Windows variations alone include Windows 3.1, Windows for Workgroups, Windows NT, and most recently, Windows 95. The first two are 16 bit vehicles; the latter use 32 bit access and deliver greater speed, stability, and flexibility. All of these run on Intel platforms which may be 80X86 or Pentium based. Fortunately, the 32 bit platforms have a Win16 module which allows a 16 bit program to run, albeit without most of the 32 bit optimization. Since a 32 bit compiler has now become available, an identical 32 bit version can be had; the cost is only the time of recompilation. Further, a substantial percentage of users prefer the Macintosh computer because of its excellent graphics and communications abilities. The Mac, too, now has two variants. These are the standard Motorola 680XX based CPU's and the new POWER Macs which use Motorola 601 - 604 RISC processors. The user survey will solicit input from users on the relative need for a version for each platform. Because the data access mechanism which we have devised is available to all of these computers and our team has the expertise to devise retrieval tools for both Mac and MS-DOS based machines, a graphical user interface will be built for both varieties of machines.

Development Tools

We evaluated three development tools which offered easy user interface design and comprehensive data access tools. These were Visual Basic, Power Builder, and Borland's Delphi. They were evaluated by the simple expedient of using them to develop some basic interface mechanisms. We chose Delphi and the reasons for this choice are detailed elsewhere.

The user interface for Windows platforms is being developed in Borland's new Delphi programming tool. This tool combines an easy to use and powerful graphical interface designer with an elementary case tool and a powerful state of the art PASCAL compiler. It produces extremely fast code. The interface keeps configuration information in a user database on the users machine. Initially, we are using Paradox, but we plan to support Microsoft Access, Visual FoxPro, and Borland's Dbase for Windows

The user interface for the Mac platform can be developed in Hypercard. It offers some of the same design features as the PC-based packages and can be easily expanded via XCMD's to accommodate features which it does not already incorporate. Unlike the Windows packages which must store information in another external database, Hypercard, itself, is also a simple database.

Communication

The user interface communicates with the proxy, which also acts as a firewall, via an Internet dialog. It does not directly connect with the Sybase Database on the R24 Server. The user interface can use most Microsoft Winsock compatible TCP-IP stacks. We have used Trumpet Winsock, SmartTerm, Microsoft Winsock (16 bit) and the new 32 bit Windows 95 version of the latter. The interface to the TCP-IP stack will be via a shareware drop-in Delphi component.

Security

Personal data which might distinctly identify any individual (Name and SSN) will not be available to the user except under an approved human subjects protocol. Security will be implemented by using a separate machine as a fire wall/Proxy server, implementing well established Internet security procedures, transmitting only PPP encoded data, and screening the user population.

Epidemiological Study

Discussions of terms, measures and procedures led to a consensus on the standardization of key research methods required to produce data comparability for joint research reporting. Common terminology was defined for first hospitalizations, unconfirmed diagnoses, age and gender stratification, case counts, crude and age-adjusted incidence rates, and person-years. Agreement was reached on the methods to be used now, at this formative stage in the collaborative process, and in the future, to classify diseases, to capture discharge diagnoses, to count cases, to calculate person-years and confidence intervals, and to age adjust. These standardized methods were used to conduct the prototype epidemiology study and are presented in **Appendix C**. This standardization is an example of the larger and very critical type of methodological cooperation required to make cross-service comparisons a reality.

The epidemiological study, itself, is described in detail in the appendix. However, it should be noted here that Air Force numerator data was derived from a SAS analysis of Standard Inpatient Data Record maintained by the Air Force Medical Support Agency; denominator data was derived from Uniform Airmen and Uniform Officer records maintained by the Armstrong Laboratory Human Resources Directorate.

Air Force, Army, and Navy enlisted women on active-duty anytime within a five-year period from January 1, 1990 to December 31, 1994 were the population used in this study. Populations were stratified by age to determine age-specific rates. This period was chosen because it encompassed a major deployment which involved women, namely, Desert Storm, and a significant post deployment period. First hospitalizations for each of eight conditions, based on hospitalization ICD-9 values, were used as indicators of disease incidence.

Results

Systems Development

User requirements

The in-house survey and external review suggested the most used variables to be included in our prototype databases. These variables included those which were used in the prototype epidemiological study also presented in this document. In-house reviewers also evaluated the content of the Navy database and recognized the value and desirability of the event code structure upon which it was built. In addition, existing databases were evaluated as candidates for potential sources of longitudinal data, remote access for joint studies, and specialized studies. Requirements for data on the Reserve and Guard components were also researched, and data sources were located and compared for all three services. Additional data requirements which would support research in other pertinent areas, such as medical readiness, were defined through interviews and in user focus groups. The results of these focus groups were presented to representatives of the Army, Navy, and Air Force at the Data Standardization Workshop held by the Air Force project team in San Antonio in December 1995. The participants used this information in determining the requirements for the Defense Medical and Epidemiological Database (DMED) standard data set. This

initial data set consists of the semantically-equivalent variables currently available from each service, which will form the data infrastructure of the planned system.

DMED Workgroup

The DMED workgroup focused on the following objectives:
- selecting epidemiologically-relevant data to be included in the DMED Data Set
- determining the level of data standardization required
- exploring data sources and defining common ground for all three services
- defining a strategy for accelerating the standardization process
- exploring a variety of technical solutions to facilitate data sharing.

The DMED Data Set was recognized as a starting point for the three services to identify and standardize data to be shared among them. This data set will provide semantically-equivalent data from all three services, which will support a variety of studies that otherwise could not be done without a great deal of time and expense. The requirements established for the initial DMED data set included the following:
- the data set would be limited to the "low hanging fruit" *i.e.*, data that are accessible to all three services and available for relatively easy inclusion at the earliest opportunity;
- the data set would focus on personnel, demographic, and medical event subsets, but would be inclusive enough to answer a variety of epidemiological questions, especially those pertaining to women's health issues;
- the data set would initially cover six years of data from 1989 to 1995;
- the data set would include only active duty members (no data on reserves unless on active duty during the period covered).

The goals of the DMED workgroup were fully compatible with the scope and direction of the Air Force Central Research Database development effort. The resulting DMED data set was incorporated into the database design for the Air Force prototype database. The DMED Data Set and a description of the proceedings of the workshop is contained in Appendix F.

Tools

The Air Force database will be constructed on an IBM R-24 which boasts a 601+ RISC processor with ample RAM and disk storage. Data has been obtained for the described six year time frame and stripped of duplicates. A six month segment has been transferred to the PC for database design and prototyping.

Data Mapping

The variables determined to be most useful by our in-house survey have been equated between Air Force, Navy, and the DoD Common Data Dictionary. Preliminary mapping tables have been constructed as Microsoft Word tables (**Appendix E**). Mapping is not always one to one and often complex. The DMED Data Standardization Workshop addressed this problem, and the participants explored ways to minimize the differences between systems. The representatives agreed that using common data sources, such as SIDR and DMDC, would significantly reduce the amount of disparity between data from different services. The Air Force and the Army Medical Surveillance System were already using similar data from DMDC and SIDR, which is in a standardized format for each service. The Navy database administrator agreed to examine these data sources for Navy data, and to use them in constructing the Navy's "partial retrofit" database for the Navy's DMED data view. The user interface software will handle the mapping and

shield the user from this complexity. The first mapping tables were used for the Joint Epidemiological study presented in the appendix.

Communications Guidelines Established

- The following guidelines were developed to allow a user to communicate with the data server.
- Communication could ONLY be effected by using the front end designed and distributed by us.
- The software could be freely distributed but unless keyed by us for a registered user could only be used in a limited fashion. The user could use some of its subsidiary tools and read on-line documents which described the system. If some data is designated as 'public' these data could be accessed. He/she could also make application to legitimately use the system.
- A user could use the software on as many machines as he wished but we would need the IP addresses of each machine.
- The software can determine the IP address of the machine on which it runs by querying Winsock. The software would convey this as part of the log-in process. If the IP address was not registered, the possessor would be invited to make an application for use of the system.
- Initially, all communication would be via the Internet and all data exchanges would be PPP encoded.
- Users would never directly communicate with the large data server but only with a small machine which would process requests and send them on to the server. This smaller machine would be responsible for checking credentials, maintaining log-in information, distributing current copies of software, and updating lookup tables. It would check queries for validity before passing them to the larger data server for processing. Since this proxy would be physically isolated from the larger machine, it could thus act as a fire wall.

Development Tools

Windows Platforms

We compared Visual Basic 3.0, PowerBuilder 4.0, and Delphi 1.0 for efficacy as development tools for Windows based clients. We decided to use Delphi because it combined the best of power and flexibility although its documentation was less than was hoped for. Delphi has been obtained and is in use with Windows 3.1/Windows NT on a 486-DX2 machine and Windows95 on a Dell Pentium 90. The latter is used as a development platform because of its speed but the interface is tested on the other machine with both Windows platforms as well

Development Tools for MacIntosh Platforms

There is no near equivalent to any of the Rapid Application Development Suites (e.g., Visual Basic, Delphi) for the MacIntosh computers. However, Hypercard, is probably the forerunner of all of these and it can now be easily expanded with XCMDs via CompileIt! and C/Pascal compilers. Because there are FTP XCMDs already available for the Mac and Hypercard itself can act as a local database, this medium will be used for both the 68XXX and Power Mac platforms. FTP services are available to the Mac via AppleScript. In addition to Hypercard's inherent database capabilities, FilemakerPro and FoxPro are being considered as alternatives for the user database. Mac development will necessarily lag behind that of the Intel-based machines because of the larger PC user base.

Software construction

The software requires development of some basic modules. These are:

An FTP client to facilitate communication with the host server system.

The user interface employs a drop-in FTP module to access the proxy FTP server. A dialog occurs between the proxy and the client as follows:

- The user interface requires the user to enter name, password, and account information. This is FTP'd to the proxy, which checks the credentials and generates a windows ini file which lists all of the Databases, Tables, and Fields to which the user has access. This ini file also includes enough information to join tables for complex queries.
- This ini file is FTP'd back to the client. The interface uses the ini file to update the local user database. Both the database and the ini file are used to dynamically present icons to the user representing databases, tables, and fields.
- After the user has constructed his query with the aid of the interface, the query, which is simply a text file, is FTP'd back to the proxy along with the ini file. The proxy checks the ini file to make sure that there is no modification, checks the users credentials again, and if all is correct, passes this text based query to the R-24 for processing.

The results of the query are passed back to the proxy where they will remain in a user folder awaiting retrieval via the included FTP client in the user interface.

The FTP dialog is carried out via an FTP client built into the interface and not modifiable or usable by the investigator. However, a true FTP client, which was developed in order to develop and test the dialog routines, will be available to the user as a module which can be used as a standard client.

All of the user interface routines have been developed. Some cosmetic changes are needed. However, the proxy machine half of the dialog has not been coded.

The login module asks the user to enter his name, password, and account information. This is checked against a file on the user's disk. If it does not match, the user is notified that he is not a recognized user and asked if he wishes to apply for use. If not, the program terminates rudely. If the user answers in the affirmative, he/she is presented with a fill-in-the-blank screen in order to enter enough information about the user to start the application process. The results of this are FTP'd up to the proxy for processing along with the user's IP address. A temporary configuration file is generated for the user which will allow the investigator to use stand alone modules such as the FTP client or access public data. The user will be sent an application form which must be filled out, signed, and returned by conventional mail.

This module has been completed except for cosmetic changes.

An interface with a local database to hold user configuration data.

The user will be able to deal with data from multiple data bases as though there was only a single entity. Standard fields, however, may have names that are cryptic to the user. The user will be able to configure each database, table, and field entity as to name and draggable icon. This information will be held in a local database which may also serve the user to hold research data. The user will also be able to set reusable constraints on data fields (*e.g.*, maxima, minima). Four popular Windows databases will be supported. These are Paradox, Access, Dbase, and FoxPro. Delphi cannot exchange graphics with Access through the available ODBC interface. Visual Basic modules can, and such modules will be shelled to for this purpose.

The basic interface to Paradox tables is complete, but some enhancements such as an icon design tool are contemplated. The Paradox interface should also work with Dbase but this still needs to be validated. The VB interface to Access tables is partially complete.

A processing engine to examine user access levels to databases, tables, and fields.

The host will FTP down a file in Windows Ini-file format. This engine will compare the new file with the current file and make appropriate changes in the local database. This is finished except for some screen design.

A 'Drag 'n Drop' interface to allow the user to graphically construct his questions.

Most of this interface design is complete and is being alpha tested by our local researchers and statisticians. Its basic functionality is as follows:

- The user must choose, in order, databases, tables, and fields. He/she is presented with a separate window for each step. The choosing sequence must be done in order but the user can go forward or backward in the sequence without difficulty.
- For each step (databases, tables, fields) the user is presented with a window divided into two panels with a section of controls and information between them.
  - ◊ The upper panel displays icons for each of the databases/tables/fields available to the user. Icons displayed are those chosen by the user. The icons are generated dynamically from the windows ini file and the user configuration database. The user configuration database is always reachable from these screens.
  - ◊ Each panel also has a list box which displays user captions for the items. Except in the case of databases, whose names MUST be unique, the caption has a numerical prefix designed to eliminate duplicate table or field captions.
  - ◊ The actual item name/table/database as well as prefixed caption is stored in the 'hint' of the image component which displays the icon.
  - ◊ When the user passes the cursor over an icon, its user caption is presented in the display bar. If help is turned on, the user also gets a description of the item in the display area.
  - ◊ A right click on the icon presents the hint information in the display bar.
  - ◊ Clicking on a user caption in the list box `highlights the icon.`
  - ◊ Double clicking on an icon brings up another floating window with a list of tables in the database or fields in a table. The user can click on a list item to see a description of the item and any notes the user has made about the item in the user database.
  - ◊ The user chooses to include a database/table/field by simply dragging its icon into the lower panel. When this happens, the prefixed caption is transferred to the lower listbox. The user can deselect by dragging the icon back into the top panel.
  - ◊ Although not implemented yet, users who are strongly text oriented will be able to drag captions from one list box to the other instead of icons. The icons will go with the text.
  - ◊ When the choices are complete, the user clicks on a button to go to the next screen in the sequence, where he/she will click a 'Show' button to show all of the tables available in the chosen databases or fields available in the chosen tables.
  - ◊ Built-in intelligence, based on information in the ini file, will help the user choose appropriate tables when tables must be joined for effective queries.
- The functionality of the database, tables, and fields screens necessarily differ. The database screen is relatively simple. User options consist only of turning help on and off. If the number of databases available becomes large, however, some of the options available in the tables and fields windows may be added.
- Tables window.
  - ◊ The user has these options
    - ♦ Turn display of item description on/off.
    - ♦ Turn table linking intelligence on/off; turn accompanying beep on/off.
    - ♦ Table linking information will always be displayed. An optional accompanying beep can be turned on/off.
    - ♦ Construct a commonly used tables (interest) group.
    - ♦ Construct a commonly chosen tables (chosen) group.

- ◆ Turn use of interest group upon show on/off.
- ◆ Use chosen group.
- ◊ Table linking intelligence works as follows: The user drags an icon to the bottom panel. If linking information is available in the ini file and the needed table has not already been chosen, the user is presented with a question such as "If you want information about **location**, you also need to include table **tbl_location**. Shall I do it for you?" The user can choose NO and go on. If the user says YES, the necessary table is transferred to the bottom panel . If this table also has linking information, help on that information will be presented in turn.
- Fields window.
  - ◊ The user has these options
    - ◆ Turn display of item description on/off.
    - ◆ Turn field join intelligence on/off; turn accompanying beep on/off.
    - ◆ Field join information will always be displayed. An optional accompanying beep can be turned on/off.
    - ◆ Construct a commonly used fields (interest) group.
    - ◆ Construct a commonly chosen fields (chosen) group.
    - ◆ Construct a repeatable parameratized query.
    - ◆ Turn use of interest group upon show on/off.
    - ◆ Use chosen group.
  - ◊ The user can constrain fields and configure figure display with these options:
    - ◆ Maxima and minima
    - ◆ Actions (*e.g.*, show, count, groupby, sum, average).
    - ◆ Complex constraint delineation
    - ◆ Field text formatting and width
    - ◆ Field captions
    - ◆ Field sorting
  - ◊ Fields which link two tables are joined by the simple expedient of dragging the icons of the two fields into a 'join' tray. A separate tray is used for each join. The user can do this manually or elect to have the built in intelligence help.
  - ◊ Fields which are to be displayed or constrained are dragged into a 'display/constrain' tray.
  - ◊ Constraints and configuration of field display is implemented by double clicking on the field icon. This opens a panel with a notebook. Pages exist for each of the sections shown above.
  - ◊ When choices are finished, the user clicks a button and the needed SQL is generated.

A translation engine which converts the users graphical query to an SQL query.

This unit is partially completed. The code to write the part of the where clause dealing with joins has been completed.

A mechanism to convert retrieved query data in the form of ASCII files into a database table.

This unit has not yet been started.

Some basic statistical tools.

This unit has not yet been started.

Epidemiological Study

Rates of occurrence of each of the eight chosen <u>disease</u> categories for each service are summarized in tables 1-8 of the joint epidemiological study presented in **Appendix D.** No outstanding differences in these rates appeared among the services.

Conclusions

Systems Development

It should be noted here that the systems team has been in place for only nine months and a considerable part of that time has been taken up with evaluating and purchasing equipment and software. Despite that drawback, we have still been able to accomplish a great deal.

Our exploration of the feasibility of producing a relational structure appropriate to research data suggests that we will have little trouble in doing so. The largest obstacle is the large amount of data processing to be done. However, data sources have been established and some of the data has arrived so that processing can begin. A Windows-based 'front end' which will seamlessly display data to the user from several databases has been started and a Mac version has been planned. A communications scheme which will allow the front end to communicate with a subsidiary machine that accesses the data server has been developed and attention has been paid to details of security during data transfer and privacy act data.

Data Server Model

It is clear from examining the needs of the Air Force and Navy research communities that a 'one size fits all' database is not the answer to the need to make research data available across the services. In addition to service specific data elements, naming and coding conventions within a service are well established and investigators are familiar with them. A better solution is to have each service maintain its own unique database but to have that database in a format amenable to SQL-queries. We envision a Data Server as able to then extract data from each database. The user would then be able to submit queries to the data server and have it return information from whatever databases are appropriate. This idea is graphically Depicted in the figure shown in **Appendix B.**

Epidemiological Study

The joint epidemiological study suggests that with appropriate cooperation, data can be shared among services if it is appropriately cross-mapped. Although most of the data used is still in non-relational format, it is extractable. The three services can communicate effectively, and a joint study such as the one given here can be produced. The availability and usability of cross service data is of great value to medical researchers from all services and will likely be equally valuable to public health specialists from the private sector as well.

One caveat which arose from the study is the acknowledgment that the inpatient data which any one service holds is only about 80% complete. This stems from hospitalizations in a treatment facility belonging to another service. The three services have agreed to exchange this data, which can now be appropriately cross-mapped.

General

Extensive cooperation and interchange has occurred between the Navy and the Air Force. An example of such cooperation is as follows.

a. The Navy has an extensive longitudinal database. It exists in flat files which are not accessible to SQL queries. However, it has data structured in an efficient manner which would lend itself to incorporation into a relational database. The Air Force does not yet have a comprehensive research longitudinal database which encompasses both demographic, personnel, and inpatient data.

b. The Air Force team has expertise in construction of relational databases and graphical user interfaces. It will develop a relational model which might be employed by both the Air Force and the Navy. The efficient event structure of the Navy database will be modified as necessary and incorporated into the relational model. The Air Force is building the user friendly front end which can query either or both databases. Both services can maintain their own conventions and the front end will equate variables between the two via mapping tables.

c. The Navy is enlarging and documenting their database; The Air Force has begun construction of theirs and the identification of source databases for the longitudinal AF data.

d. Meanwhile, although the databases and data server are in their infancy, researchers and systems experts from both services collaborated to produce the prototype epidemiological study presented here, even though the expert tools were not yet in place.

e. In December 1995, representatives of the Army, Navy, and Air Force met in San Antonio to explore technical solutions to facilitate the sharing of semantically-equivalent data across the services. During these meetings, the group developed the DMED standard data set, which represents the common data currently available and deemed most useful for cross-service epidemiological studies. These data will be further analyzed and mapped to DoD standard data elements to produce true semantic equivalence across the services. The Air Force has incorporated these data into its database design to provide a "DMED data view". The Navy will create a relational database using this data set as the foundation, and the Army plans to make this data view available from their Medical Surveillance Database.

f. Although more work must be done for the technical implementation of the communications infrastructure, this collaborative project between small Air Force and Navy teams resulted in tremendous progress in a very short time. The development teams accumulated a great deal of knowledge about the current systems, user requirements, data standards, and technical solutions. The Air Force team is continuing in the direction begun during this project, and will coordinate with the other services to ensure that the common communications infrastructure soon becomes a reality.

Everyone in the medical research community benefits from such cooperation. Researchers interested in Women's health benefit especially because the relative number of women in the services is small. When divided into population groups, accurate statistics may be difficult. With two or more populations to work with, more information may be extracted from the data.

Appendices

Appendix A - Database Primer

The following overview of the relational database management system is intended to assist the reader in understanding the concepts presented in the Air Force team's technical approach. This brief introduction describes a few of the different types of file management systems, and includes definitions of some of the common terminology. It does not attempt to explain how to design a database, choose the appropriate database management system, or manage the project. This discussion does try to illustrate some of the many steps and processes involved in developing a database which is technically feasible, efficiently designed, and meets its users' requirements.

The "Flat File". A *database* is simply a collection of data, organized into at least one file, that is fundamental to a system. The simplest type of file management system centers around the *sequential file*. This structure requires that the file be read (or written) one record at a time, where each record follows its predecessor physically on the storage device. The data records may be sorted in some sequence, although that is not necessary. Normally, the entire file is "passed" whenever processed. The Navy's longitudinal enlisted history database, in which most of the data are stored in one large COBOL flat file, is an example of this type of system.

The advantages associated with the "flat file" structure are that it is easy to create and simple to use. It also requires minimal overhead to access and use. Some of the underlined disadvantages of the "flat file" structure include the following:
- Any change in structure or content require simultaneous changes to all programs
- Relationships across files must be imbedded within application logic
- The structure encourages redundant data to ease the overhead of gaining access to data when needed.

The Database Management System. A database management system (DBMS) consists of the database file structures and software which uses various methods to interact with the database to maintain data, update the database, and extract data for queries and reports. There are three classic types of database management systems: the *hierarchical system*, the *network system*, and the *relational database system*. All three systems offer some of the same advantages over the flat file system:
- The DBMS supports data independence (application programs do not need to be concerned with the physical implementation - that is the job of the DBMS)
- The DBMS can support complex data relationships
- The DBMS provides enhanced security mechanisms for access to data
- The DBMS provides sophisticated backup and recovery mechanisms
- The DBMS provides advanced features, such as on-line and ad hoc access, which greatly enhance the system's value and usability.

The Relational Database. The most flexible of these types of database management systems is the relational database management system, or RDBMS. A *relational database* can be thought of as a collection of flat files, or tables. There are no linkage paths or structural connections between associated records. The relationships between records are based on the *data content of the records themselves*. Without the "hard-wiring" associated with other types of DBMS, the relational database structures can be easily modified to meet new or changing requirements. For example, a

patient may be admitted to a hospital with a number of different diagnoses. In a relational database, the patient's demographic data (date of birth, SSN, sex, marital status, etc.) could be stored as a record in a "patient data" file, and his diagnoses for that admission episode could be stored in a "diagnosis" file. The data would be **related** by a common field in both files, such as the patient's SSN. Thus, additional diagnoses for the patient could be entered as records in the diagnosis file. There are some definite advantages in organizing data this way. For example, a query or report requiring only patient data would not have to access the diagnosis file. This design makes it easier and more efficient to retrieve the desired data, as well as to maintain the database. A few of the popular relational databases today include Sybase, Oracle, Informix, Access, and Paradox. Sybase can handle massive databases, while FoxPro, Access, and many other popular systems are designed for the PC. A big advantage of the RDBMS is the ease with which queries can be formulated and data can be retrieved. **SQL**, "Standard Query Language", is widely recognized as the standard language for relational databases. SQL makes it possible to formulate queries to retrieve data from different relational databases at different locations. Many relational database systems include communications capabilities for accessing data from remote sites. ODBC, or "Open Database Connectivity" is a standard which enables this remote communication between relational databases. Thus, the concept of a "virtual database" is now a reality; a user can retrieve and merge data from relational databases in different locations to create unique data sets on his personal computer.

The Data Model. The process of analyzing and organizing data in a way in which it can be translated into a relational design involves quite a few steps. Trying to rush through this process can result in a very inefficient database design, and one which does not meet the users' requirements. Another very important task, which can greatly impact the whole project, is to properly define the system's scope with its users. If the scope is large, and if the approach involves using new technology, it is wise to first develop a **prototype** system. The prototype will demonstrate whether the technical approach is feasible, and will provide the development team with important user feedback to guide future design efforts.

Here are some of the tasks involved in this process before the actual database development work begins:

- identify the users who will be using the database, and for what purposes
- work closely with the users to define their data requirements, and create clear and usable definitions of the data
- model the data requirements by organizing the data into **entities** with their associated **attributes**, and by establishing relationships among the entities. (see definitions for these terms in the section following)
- work through a review process called "normalization" to eliminate redundant or mutually exclusive data, consolidate synonyms, clarify definitions, and generally, to make sure that every piece of data is in its proper place in the model.
- evaluate the users' data usage requirements (i.e., which data do they require, how often, how should the data be sorted, how complex are the queries, how many records are involved, etc.)
- apply the users' data usage requirements to the data model to define data "usage paths". The database designer can use record indexing and other features offered by the RDBMS to facilitate rapid sorting and retrieval of most frequently demanded data.
- review the data model with the users to make sure it reflects their requirements.

At this point, it is sometimes possible to begin constructing the relational database design right from the data model. However, the process of developing the data model should not be rushed; in this case, careful attention to the data model pays off in a more efficient database design, and one which will meet its users' requirements.

Definitions. The following terms are used in this discussion. Some terms are interchangeable, so you can relate to the term you are most comfortable in using. Some of the terms have more of a "data model" connotation, in that they are used to define the user's data requirements. Other terms are more familiar to information systems personnel in the systems development process. These definitions offer simple examples to illustrate the data modeling and database design processes.

Entity. An entity is something about which information is known. It describes or represents something, such as a person, a place, or a thing. Examples of entities include patient, active duty member, training program, military treatment facility. Entities are identified and described by users and systems analysts working together to define the data requirements for the system.

Attribute. An attribute provides detailed information about an entity. It helps identify, describe, or clarify an entity by providing a value for a quantifiable characteristic or trait. For example, an active duty member has a name, SSN, date of birth, and pay grade. Two or more entities can be related by having a common attribute (i.e., SSN). The entities, their attributes, and the relationships between the entities form a large part of the **data model** for the system.

Normalization. The process of reviewing the entities, attributes, and relationships in a common sense approach, to eliminate redundant data, resolve one-to-many data relationships, consolidate mutually exclusive data, resolve other conflicts in naming and definitions, and to make sure that each attribute is assigned to the entity to which it logically belongs. This process streamlines and cleans up the data model, and provides a logical resolution to the relationships between entities. During this process, **keys** are assigned to define the relationships between the entities.

Primary key. An attribute or combination of attributes that uniquely identifies one occurrence of an entity. This key must be contained in each record for that entity. (i.e., a combination of name and SSN)

Foreign key. An attribute (or combination) that appears totally as a primary key in another relation. This key relates two or more entities. The entity "person" may have one or more addresses; in a RDBMS, this relation may be shown by the person's ID (primary key) appearing as a foreign key in an address record (occurrence of the entity "address")

Data Dictionary. A collection of information about the entities and their attributes in the data model, including definitions, aliases, parameters, ranges, formats, rules, references, and other significant data.

Relation. A two-dimensional array or table of data containing descriptive information about an entity. ("relation", "file", and "table" are used interchangeably)

Column. An attribute (variable) defined within a specific table. ("column", "attribute", "field", and "variable" will be used interchangeably)

Row. One occurrence of a record within a table. ("row" and "record" are used interchangeably); a row consists of one occurrence of each column, or attribute, in the table.

ODBC. Open Database Connectivity is a standard which enables communications between relational databases.

SQL. Standard Query Language, the universal language of relational databases.

Examples of Relational Tables. In the following table, the columns show attributes, or variables associated with the entity "patient", and the rows show the occurrences of each record in the table.

Notice that the records and columns do not appear in any particular order. This is not important in a RDBMS; the system is so flexible that changes to the database may be made fairly easily.

Patient Name	Register Number	Patient Category	SSN	Sex
Cable, Samuel M.	3256774	F11	111220000	M
Barger, Margaret	3256771	F11	222334444	F
Miller, Sandra	3256773	N11	333669999	F
Popper, James	3256772	F14	111009992	M

In the next table, a patient can have multiple diagnoses for a single admission episode. The **key** which relates each diagnosis back to the right patient and admission episode here consists of register number and SSN. In this case, register number identifies an admission episode. One could argue that register number could uniquely identify a patient in a single admission episode, but there may be cases in which there are duplicate register numbers accidentally assigned (as when assigned by hand, when the system has gone down, and data must be entered later.) This is an example of an event that must be taken into consideration when designing the database, and when determining which attributes uniquely identify a single occurrence of the entity.

SSN	Register Number	Diagnosis (ICD-9-CM code)	Diagnosis Number
111009992	3256772	403.9	1
111009992	3256772	372.03	2
222334444	3256771	371.57	1
111220000	3256774	416.9	1
333669999	3256773	518.4	1

Appendix B - Data Server Model

Air Force Central Research Database and
Central Data Server
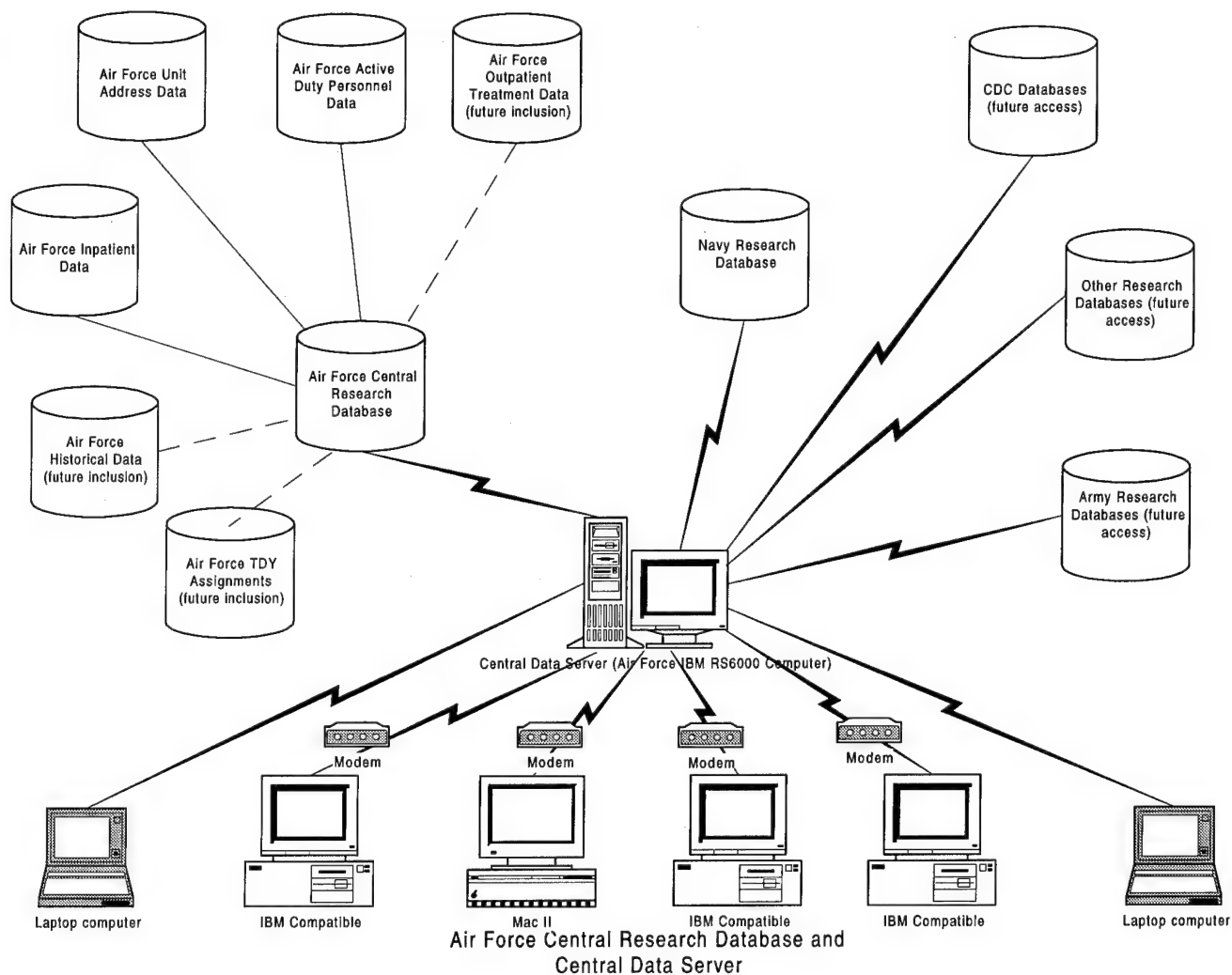
Appendix C - Statistical Methods and Definitions

Table 1. Standardized Methods for Cooperative Multi-Service Research

Term/ Procedure	Description	Standardized Measures
Age Adjustment	A procedure for adjusting rates, designed to minimize the effects of differences in age composition when comparing rates for different populations.	Age adjustment is performed using the indirect method of standardization (see Standardization).
Age-Adjusted Incidence Rate	The result of age adjusting: A rate that controls for the age differences between populations.	A summary rate using indirect standardization to age adjust. These are the rates that would be expected if the Air Force, Army and Navy all had the same age distribution.
Age Groups	Stratification into several subgroups can be used to control for the effect of confounding variables such as age. Without such a mechanism, one can mistake the effect of differences in age composition for differences in disease rates, when in fact the age-specific rates are equivalent.	Age groups are defined as follows: 17-19, 20-21, 22-24, 25-29, 30-34, 35-39, 40-44, and 45 and above. The 17-19, 20-21, and 22-24 year-old groups represent more finite categories than the older age categories groups. This was done because of the large number of young personnel in the military.
Case	Individuals identified as having the particular disease, health disorder, or condition under investigation.	The number of first hospitalizations for all individuals observed during the study period, per diagnosis. For this study, services are reporting cases of their own personnel admitted to their own hospitals and not to other service hospitals. For future studies, each service will exchange hospitalization data for those persons admitted to their service-specific hospitals who belong to the other services.

Table 1. Standardized Methods for Cooperative Multi-Service Research (continued)

Term/ Procedure	Description	Standardized Measures
Confidence Interval (95%)	Since the sample may not be representative of the larger population if the data were complete, or of a longer time period of observation, a statistical estimation is made to determine the range or interval of values that has a 95% probability of including the true incidence rate. The interval is bounded by an upper and lower confidence limit.	Confidence limits (intervals) around the incidence rates using the Poisson distribution for rare occurrences are employed. (Haenszel)
Crude Incidence Rate	A measure of the frequency with which a particular event occurs in a defined population. All rates are ratios, calculated by dividing a numerator, (e.g., the number of hospitalizations for kidney disease), by a denominator, (e.g., the total population at risk of contracting kidney disease for a specified time period). The population denominator can be expressed as the total number of people at risk or as person-time units (e.g., person-years) at risk.	The number of first hospitalizations per 100,000 person-years, without adjusting for age.
Denominator	The lower portion of a fraction, (e.g., the population at risk), used to calculate a rate. The denominator can be expressed as the average population at risk during a specified period, or as person-time units at risk for that period. (see person-years)	The population at risk expressed in person-years.
Diagnosis	The identity of the condition from which a patient suffers.	The International Classification of Diseases, 9th Revision (ICD-9) codes was used to classify diagnosis.

Table 1. Standardized Methods for Cooperative Multi-Service Research (continued)

Term/ Procedure	Description	Standardized Measures
Diagnosis (continued)		The Army stores data on a maximum of 8 discharge diagnoses (primary, secondary, etc.), the Navy 10, and the Air Force 20. The Army searches 8 deep to identify diagnoses of interest, the Navy searches 10 deep, and the Air Force searches through only the first 10 of its 20 hierarchical discharge diagnoses.
First Hospitalization	The first inpatient admission within the specified interval for a unique ICD-9 diagnosis. Second and later hospitalizations for the same diagnosis are excluded in order to provide unduplicated counts.	For this study, only the first hospitalizations for the full 5-year interval were used. For future studies, subsequent hospitalizations may be included based on specific criteria (to be determined), as appropriate for certain diagnoses.
Numerator	The upper portion of a fraction used to calculate a rate. The number of cases or occurrences of first hospitalizations.	Occurrences of first hospitalizations or cases.
Person-Years at Risk	A more accurate way to calculate rates when the population is not constant and people enter or withdraw at different points during the period of observation. It is the sum of the individual years that persons in the study population have been at risk of contracting the condition of interest. Each person contributes only as many years as he/she is actually observed in the time interval of interest, i.e., 100 person-years may represent 100 persons for one year, 1 person for 100 years or any intermediate mixture between these extremes.	This denominator combining persons and time is the sum of the number of days each person at risk contributed during the observation period, divided by 365.25.

Table 1. Standardized Methods for Cooperative Multi-Service Research (continued)

Term/ Procedure	Description	Standardized Measures
Standardization	A set of techniques used to remove as far as possible the effects of differences in age, gender or other confounding factors, when comparing two or more populations. For purposes of this study, "standardization" is synonymous with age-adjustment. Two methods of standardization can be used, direct or indirect. The most frequently adopted method is the indirect method. It is used to compare study populations when some age groups have small sample sizes producing specific rates that may be too imprecise and unreliable, i.e., statistically unstable, for use in detailed comparisons.	Since the population under study may be statistically unstable, indirect standardization as a method of age adjusting is used. Indirect standardization involves a series of mathematical steps as described by Lilienfeld. The specific rates in the standard population are averaged, using as weights the distribution of the study population. The ratio of the crude rate for the study population to the weighted average obtained is the standardized incidence ratio or SIR. The indirect standardized rate is the product of the SIR and the crude rate of the standard population (Note: a SIR is the incidence equivalent of a Standardized Mortality Ratio or SMR; see Lilienfeld for an explanation of SMR's). (Lilienfeld)
Standard Population	A commonly available population of known age distribution.	For this report the sum of the person-years for Air Force, Army, and Navy enlisted women between 1990-1994 was used. For future studies, a total force standard population for an as yet to be determined time interval remains to be selected.

Appendix D - Joint Epidemiological Study

Cooperative Prototype Tri-service Age-specific Study of Selected Disease Incidence in Military Women Using Inpatient Hospitalization Data

Introduction

Congress has mandated a medical research program to focus on the specialized health needs of military women, including patterns of illness and injury, environmental and occupational hazards, and psychological stress, as well as the development of a database to facilitate long-term research studies (1). The number of women in the military is growing. Currently, active duty women in the Armed Forces comprise approximately 14% of the total population. In upcoming years the percentage is projected to reach 20% (2). The expanding population of women in the military offers a distinctive and reliable number of potential subjects for long-term, gender-specific studies. It can also provide the military with valuable clinical data for both military and civilian women (1). Comprehensive incidence rates from such studies would be useful in planning preventive strategies, and in providing adequate and appropriately trained treatment personnel and facilities for women.

Epidemiological studies in the military have historically been limited to a service-specific data source, consequently no previous study has used joint multi-service data to identify the incidence of hospitalizations. Furthermore, no multi-service study has been conducted that has focused exclusively on women although several studies have examined gender differences in hospitalization patterns, but they have been confined to a specific service (3,4). In one study, pregnancy-related conditions, primarily the delivery of newborns, were determined to be the leading reason for hospital admissions (33.7%) in women (3). There are some diseases, however, such as endometriosis and pelvic inflammatory disease that are unique to women (5,6). Also, there are conditions that can occur in both sexes which disproportionately affect women such as kidney disease (6), systemic lupus

erythematosus (6,7), and genitourinary disorders (8). Other conditions are unique to or disproportionately affect men. Examples include prostate cancer (5), testicular cancer (9), acts of violence (5), alcohol dependency (6), and accidents and injuries (10-11).

The purpose of the present study was to provide standardized methods for calculating epidemiological measures which would become the model for future multi-service women's health research. These methods were used to assess and compare disease incidence in active-duty enlisted Air Force, Army, and Navy women. First hospitalization rates were used as a measure of disease incidence. Eight disease categories were selected for investigation. The diseases selected were believed to be unique to or more common in women, possibly amenable to preventive intervention during military service, and of special interest to the women's health research community.

Methods

Computerized records of Air Force, Army, and Navy enlisted women on active- duty anytime within a five-year period from January 1, 1990 to December 31, 1994 were searched for hospital discharge diagnoses. Each service used its own resources to retrieve this data and to acquire its respective population totals. The Air Force obtained numerator data from the Standard Inpatient Data Record maintained by the Air Force Medical Support Agency, and denominator data from the Uniform Airmen and Uniform Officer records maintained by the Armstrong Laboratory Human Resources Directorate. The Army used its already existing surveillance database, the Army Medical Surveillance System to obtain data. The Navy data was obtained from a career history data system containing comprehensive information on service history and hospitalizations for all Navy personnel (9, 14, 33-38). The Navy named its system the Career History Archival Medical and Personnel System (CHAMPS).

Population

Air Force, Army, and Navy enlisted women on active-duty anytime within a five-year period from January 1, 1990 to December 31, 1994 were the population used in this study. Populations were stratified by age to determine age-specific rates. Age groups were defined as follows: 17-19, 20-21, 22-24, 25-29, 30-34, 35-39, 40-44, and 45 and above. The age span of these groups are not equivalent because of the large number of young personnel in the military. Use of larger intervals for younger personnel would cause valuable information to be lost.

Numerator Data

The three services decided to use first hospitalization as a measure of disease incidence for eight selected diagnoses. First hospitalization for each of eight diagnoses was defined as the first inpatient admission between 1990 and 1994 for a unique International Classification of Diseases, 9th Revision (ICD-9) code (12). The eight selected diagnoses were: personality disorders (301.0-301.9); adjustment reactions and disorders (depressive) (309.0-309.1); adjustment reactions and disorders (other) (309.2-309.9); cholelithiasis (574.0-574.5); cholecystitis (575.0-575.1); kidney infections (590.0-590.9); pelvic inflammatory disease (614.0-614.9); and endometriosis (617.0-617.9). Second and later hospitalizations for the same diagnosis for the same person were excluded in order to provide unduplicated counts. Thus, one individual might have been hospitalized more than once for the same condition, but only her first admission between 1990-1994 was included.

Denominator Data

Two types of population data were obtained: the Army and Navy used the period of enlistment for each enlistee to calculate actual person-years and the Air Force used end-of-year annual strength totals to approximate person-years. The total number of person-years during this five year period for the Air Force was 280,001, for the Army, 308,797, and for the Navy, 217,873. The Navy and the

Army defined person-years rate as an incidence rate that directly incorporates time into the denominator. It was calculated by dividing the total number of cases during the observation period by the time (in days) each person was observed, multiplied by 100,000 (13). The rates were expressed as the number of first hospitalizations per 100,000 person years.

Statistical Procedures

Confidence intervals around the crude incidence rates were calculated using the Poisson distribution, for rarely occurring events. The Poisson distribution was used because the disease incidence rates of the three services had many rates ranging from 1 to 100 per 100,000 population; sample sizes of 10,000 to 1,000,000 would be necessary before the normal distribution could safely be used to calculate confidence limits (13). A table of 95 percent confidence intervals developed by Haenszel, Loveland, and Sirken was used to calculate the 95 percent confidence limits for low incidence rates exemplified in this study (13). Because these are based on rarely occurring events, the interval was wider than that of a normally distributed confidence interval, reflecting the uncertainties based on estimates involving a small number of cases (13).

Nearly all comparisons of rates among populations require statistical adjustment for age or other factors known to influence the distribution of the disease under study. Appropriate methods for assessing the statistical significance of observed differences must also be used (14). In this study, age-specific rates were used to calculate age-adjusted service-specific total incidence rates, using all three services as a standard population. Adjustment is often performed on rates because of differing age distributions in populations that are being compared (15). Because age can have a marked effect on morbidity and mortality, statistical procedures were conducted in this study to "remove the effect" of age within each service (16). The procedure used in this study to adjust the rates for age differences in the three services was the indirect standardization method because some age strata were comprised of small numbers of women, and the associated specific rates were too imprecise for use in detailed comparisons among the three services (17). Table 1 lists multi-service standardized methods for all components of this study in detail.

Results

Age-adjusted incidence rates varied among the three services depending on the particular ICD-9 diagnosis category. The rates for females range from a high of 1141.3 per 100,000 person-years for pelvic inflammatory disease (Army) to a low of 27.2 for cholecystitis (Air Force). Age-specific incidence rates for women in all services (Air Force, Army, and Navy) are shown in Tables 1-8. These tables are preliminary and represent data that may be incomplete and remains to be verified.

Psychological Disorders:

Adjustment Disorders, depressive (ICD-9 codes 309.0-309.1) and other (ICD-9 codes 309.2-309.9); Personality Disorders (ICD-9 codes 301.0-301.9)

Adjustment disorders are defined in the DSM-IV (18) as "the development of clinically significant emotional or behavioral symptoms in response to an identifiable psychosocial stressor or stressors". A common adjustment disorder is depression, which is manifest by symptoms such as depressed mood, tearfulness or feeling of hopelessness. Other adjustment disorders include anxiety, disturbance of conduct (e.g., truancy, vandalism, reckless driving, fighting, defaulting on legal responsibilities) and maladaptive reactions (e.g., physical complaints, social withdrawal, or work or academic inhibition). Adjustment disorders can be acute (less than 6-month duration) or chronic (6 months or longer).

Personality disorders are defined in the DSM-IV (18) as "an enduring pattern of inner experience and behavior that deviates markedly from the expectations of the individual's cultural, is pervasive and inflexible, has an onset in adolescence or early adulthood, is stable over time, and leads to distress or impairment". Specific personality disorders include paranoid, schizoid, schizotypal, antisocial,

borderline, histrionic, narcissistic, avoidant, dependent, and obsessive-compulsive personality disorder.

Rates of adjustment disorders vary depending on the population studied and the assessment methods used. In outpatient mental health treatment, the percentage of individuals with a principal diagnosis of adjustment disorder ranges from 5% to 20% with little overall gender variation (18). While many studies suggest that the prevalence of depressive disorders is more common among women than men, (19-21) there remains considerable controversy as to whether this gender difference is true. (22)

Certain personality disorders (e.g., antisocial personality disorder) are diagnosed more frequently in men, while others (e.g., borderline, histrionic and dependent personality) are diagnosed more frequently in women. (18) A recent study confirmed a greater prevalence of antisocial and narcissistic personality disorders in men, but failed to identify a higher prevalence of any personality disorder in women. (23)

Crude and age-adjusted first hospitalization rates for adjustment and personality disorders among Air Force, Army and Navy personnel are summarized in Tables 1-3 for the years 1990 through 1994. Adjustment disorders are divided into depressive disorders (ICD-9 codes 309.0-309.1) and other disorders (ICD-9 codes 309.2-309.9). Personality disorders include ICD-9 code 301.

The first hospitalization incidence rates for adjustment disorders (depressive) are shown in Table 1. The total rates ranged from a low of 195.0 per 100,000 person-years (95% CI, 178.4-213.1) in the Air Force to a high of 416.4 in the Army (95% CI, 391.0-443.1). Air Force and Navy rates were lower than Army rates for all age groups. Air Force and Army rates decreased steadily until age 35-39 when the rates increased. Except for a slight rise in 30-34 year-olds, Navy rates also decreased steadily through age 61.

The first hospitalization incidence rates for adjustment disorders (other) are shown in Table 2. The total rates ranged from a low of 304.9 per 100,000 person-years (95% CI, 283.0-328.7) in the Navy to a high of 569.3 in the Army (95% CI,

534.6-605.7). The highest crude rate was noted in Army women ages 17-19 who had a rate of 1546.5. All three services' rates decreased steadily until age 40-44 where the rates for all services increased, but then decreased again in the 45-61 age group.

The first hospitalization incidence rates for personality disorders are shown in Table 3. The total rates ranged from a low of 225.3 per 100,000 person-years (95% CI, 207.7-244.2) in the Air Force to a high of 781.6 (95% CI, 733.9-831.6) in the Navy. The age group at highest risk was the 17 to 19-year-old group for all three services. Similar to adjustment disorders, the incidence of hospitalization for personality disorders also declined with age.

Kidney infections (Pyelonephritis)

One complication of urinary tract infections is pyelonephritis. Acute uncomplicated urinary tract infection is one of the most common problems for which young women seek medical attention and accounts for considerable morbidity and health care costs (24). It has been estimated that 25 to 35% of women between the ages of 20 and 40 years have had at least one urinary tract infection (25) and the prevalence of urinary tract infection increases with age, reaching almost 7.0 per cent in women over the age of 50 years (26).

Acute cystitis is a superficial infection of the bladder mucosa, whereas pylonephritis involves tissue invasion of the upper urinary tract. As many as one-third of all episodes of acute cystitis are associated with silent upper tract involvement (24). Recurrent urinary tract infections, however, do not lead to permanent kidney damage or end stage renal disease (27).

Sexual intercourse, and use of a diaphragm and spermacide are the two behavioral factors that have been most consistently associated with urinary tract infections among women (27). Biological causes of urinary tract infection have focused on the interplay between minor weaknesses in host defenses and bacterial virulence factors (27). Most urinary tract infections are caused by *Eschericia coli* (E. coli) (24), although one recent study reported that *Staphylococcus*

saprophyticus caused 13% of upper urinary tract infections (28). P-fimbriae, mannose resistant haemagglutination and the production of haemolysis have been associated with strains of E. coli isolated from patients with acute pyelonephritis (29).

The expression and specificity of urovirulence determinants, particularly fimbrial adhesins, hemolysin, and aerobactin among *E. coli* isolates recovered from women with urinary tract infections have been studied (see review, 28). Fimbrial adhesins mediate attachment to uroepithelial-cell, receptor molecules. Adherence or attachment of uropathogenic E. coli to the mucous membranes of the genitourinary tract is necessary for virulence and the production of acute disease (28). Women with recurrent urinary tract infection may have more adhesin receptors on their genitourinary mucosa and therefore more binding sites for E. coli since uroepithelial cells from infection-prone women bind more E. coli per cell on average than do cells from healthy women) (28). Adhesins are also the main invasion mechanism in men, but it has been suggested that males are significantly more resistant to urinary infections both in the tract and parenchyma than women (29).

Many women experience recurrent infections. Brauner et al (30) in a 38-month prospective study of 23 women with acute pyelonephritis due to E. coli, reported that despite treatment and repeated negative urine cultures, each woman had 1-4 new episodes of E. coli bacteria caused by E. coli strains identical to the one that caused the initial episode. Stamm et al (31) observed 51 infection prone women for eight years and reported that when these women were not receiving antimicrobial prophylaxis, infections occurred at an average rate of 2.6 per patient year, but varied widely from patient to patient. These results suggest that the infecting E. coli strain may survive in fecal flora or is harbored in patients' surroundings (30).

There have been few studies comparing the rates of pyelonephritis and other urinary tract infections among men and women. Recent statistics for inpatients

discharged from short stay, nonfederal hospitals in 1991 with the ICD-9 code 590 indicate a female to male ratio of 4.1:1 (32).

The first hospitalization incidence rates for kidney infections are shown in Table 4. The total rates ranged from a low of 197.3 per 100,000 person-years (95% CI, 179.7-216.6) in the Navy to a high of 354.0 in the Army (95% CI, 332.4-376.7). Within the Army kidney infection rates decreased steadily with age, except in the 45-61-year-old age group where the rate again rose. Kidney infection rates among Navy women showed no consistent linear pattern, but were generally lower among older age groups. Air Force kidney infection rates generally decreased with age, but rose again in older age groups. A high incidence of pyelonephritis and urinary tract infections could lead to increased heath care costs and time lost from work, especially among female military personnel. Further research is needed to examine differences among women by branch of military service, by deployment status, and by paygrade.

Incidence rates of five other diagnoses were investigated in this study. The rates for cholelithiasis are shown in Table 5. They were similar among the three services for all age groups, and generally increased with age. Rates for cholecystitis are shown in Table 6. They appear comparable among the three services until ages 35-39 when Navy incidence rates increase. Rates for endometriosis are shown in Table 7. Rates increased with age, confirming that women in older age groups are at higher risk . In fact, endometriosis occurs most commonly in women between ages 30-40 (6). Finally, rates for pelvic inflammatory disease (PID) are shown in Table 8. The age-adjusted rates for PID are the highest of any diagnosis observed across all three services. In the Navy, crude rates increased steadily, peaking at 1738.8 per 100,000 person-years in the 40-44 year-old age group, but declined to 481.4 in the 45-61 year-old age group. Generally, the same pattern was seen in the Air Force and Army, with higher rates in the 25-44 year-old age groups.

Discussion

This study was undertaken to demonstrate the feasibility of conducting cooperative multi-service research into women's health and to develop standard methods for future medical research projects. The primary intent was to identify research-related issues, to define common terminology and objectives, and to develop standardized methods for conducting joint research. The data produced in carrying out this demonstration are preliminary and incomplete in some cases. Therefore, these data are not suitable for drawing inferences. After the services have gained more collective experience refining these research standards and procedures, the focus can shift to the actual research outcomes, to statistical inferences, and to the implications for policy, prevention and treatment of servicewomen.

In this exploratory investigation of first hospitalization rates among Air Force, Army, and Navy enlisted women eight diagnoses were examined. The results have provided preliminary data on the differences in hospitalization rates among women of different ages and between the services and a profile of those age-specific populations which appear to be at increased risk for these diseases. Statistical testing to compare age-specific and service-specific risk was not an objective of this study, but could be undertaken in the future.

Several limitations of these results, either due to study design or to artifact, must be addressed when interpreting the results. Artefactual limitations are mechanical or procedural in nature and can result from between service differences in whether, how and when cases are counted. These service differentials encompass: (1) obtaining data, i.e., the amount of missing or incomplete data, (2) how cases are counted, (3) whether women with particular conditions are hospitalized or treated in outpatient or rehabilitation centers, (4) certain conditions which may be differentially associated with military discharge before a first hospitalization occurs, and (5) newly developed standardized methods and data management processes that require optimization. Artefactual limitations could include for example, differences in criteria and how they were

employed to determine who should be hospitalized, differences in how medical records are kept, differential delays in obtaining hospital data and person-year (length of service) data, and different procedures between the three services in terms of what criteria were used to determine who should be hospitalized.

First hospitalizations may not provide a complete picture of incidence rates for some of these eight diagnoses since only inpatient admissions were included in this study. Outpatient visits and the large number of cases that never come to the attention of the health care system were not included. Incidence rates of first hospitalizations therefore represent only the most severe cases, the proverbial "tip of the iceberg." True incidence rates of these diseases may, therefore, be significantly under-reported. Furthermore, the degree of under-reporting is not uniform for all eight diseases investigated, because some conditions such as cholecystitis are very likely to result in hospitalization (6), while psychological conditions for example, in general are treated in ambulatory facilities. Also there were differential delays in obtaining hospitalization data across the services and in some cases, data were not available for the latter part of 1994. Finally, there may be service-specific factors such as serving aboard ship or flying an aircraft that differentially affects who gets hospitalized.

There are also epidemiological study design limitations to this study. Study design limitations include issues of stratification and adjustment for other factors to control for confounding, and the possibility of service differences in the utilization of hospital treatment by military women. Also, the hospitalization data was not stratified by or adjusted for such potentially important confounders as race, occupation (different occupations pose different risks for varying diagnoses) or markers of socioeconomic status (SES). Important differences in the racial, occupational or SES composition of the three services may account for some of the rate differences presented. Future studies can be designed to address this limitation.

Despite these limitations, incidence rates based on first hospitalization events can accurately characterize the most severe spectrum of illnesses and injuries

affecting a population. Many previous studies have used incidence rates based on first hospitalization events (3, 9, 33-38). The ability of the Department of Defense to identify and characterize the most serious medical problems facing active-duty servicewomen will be enhanced by collaborative epidemiologic research exemplified by this prototype study. The results of future studies modeled after this prototype study have several implications for Air Force, Army, and Navy health care providers. For example, they may help drive preventive medicine efforts throughout the military. They may also demonstrate that a better understanding of specific diagnoses and their associated risk factors can be beneficial in preventing and controlling these diseases. Also, these rates can provide the basis for efficient planning and allocation of medical personnel and resources to assist in maintaining military readiness. Finally, results of future research may be generalizable to all women and significantly improve women's health, medical care and quality of life. By conducting future studies of women's health, a wealth of medical data lies waiting for multi-service exploration.

PRELIMINARY FINDINGS TABLES:

P R E L I M I N A R Y   F I N D I N G S

Table 1. Incidence rates of first hospitalizations for adjustment reactions and disorders (depressive) (ICD-9 309.0-309.1) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

	Air Force enlisted					Army enlisted					Navy enlisted				
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper
17-19	116	22496	515.7	428.0	620.8	236	26964	875.2	759.7	1007.4	117	24495	477.7	391.1	564.0
20-21	100	41171	242.9	198.7	296.3	280	52044	538.0	474.5	610.1	110	42167	260.9	212.1	309.5
22-24	118	56764	207.9	172.9	249.9	261	65802	396.6	349.8	449.7	83	53648	154.7	121.4	187.9
25-29	107	67750	157.9	130.0	191.6	268	74195	361.2	318.6	409.6	80	53679	149.0	116.4	181.6
30-34	40	47644	84.0	59.9	114.2	120	51040	235.1	195.8	282.1	46	27818	165.4	117.6	213.0
35-39	32	32769	97.7	66.8	138.1	77	27890	276.1	216.7	350.6	19	11958	158.9	87.6	229.8
40-44	11	10322	106.6	53.2	190.8	24	8439	284.4	182.3	423.8	4	3278	122.0	4.6	238.8
45-61	1	1085	92.2	2.3	513.4	9	2423	371.5	170.1	705.9	1	831	120.4	0.0	240.7
Total	525	280001	187.5	171.9	204.5	1275	308797	412.9	387.7	439.3	460	217873	211.1	191.8	230.4
Age-adjusted rate **			195.0	178.4	213.1			416.4	391.0	443.0			199.0	181.3	218.5

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

P R E L I M I N A R Y   F I N D I N G S

Table 2. Incidence rates of first hospitalizations for adjustment reactions and disorders (other) (ICD-9 309.2-309.9) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

	Air Force enlisted					Army enlisted					Navy enlisted				
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper
17-19	270	22496	1200.2	1063.4	1354.8	417	26964	1546.5	1401.1	1707.3	164	24495	669.5	567.0	771.8
20-21	190	41171	461.5	399.2	533.3	422	52044	810.9	734.7	895.2	189	42167	448.2	384.3	512.0
22-24	156	56764	274.8	234.2	322.5	333	65802	506.1	451.4	567.3	138	53648	257.2	214.3	300.1
25-29	154	67750	227.3	193.4	267.1	289	74195	389.5	343.5	441.7	131	53679	244.0	202.2	285.8
30-34	67	47644	140.6	109.8	179.9	164	51040	321.3	274.4	376.2	68	27818	244.5	186.3	302.4
35-39	35	32769	106.8	74.5	148.5	79	27890	283.3	222.4	359.8	21	11958	175.6	100.6	250.2
40-44	19	10322	184.1	110.8	287.2	30	8439	355.5	240.0	508.4	7	3278	213.5	56.7	369.4
45-61	0	1085	0.0	0.0	0.0	3	2423	123.8	25.5	361.5	0	831	0.0	0.0	0.0
Total	891	280001	318.2	297.7	340.0	1737	308797	562.5	528.2	598.5	718	217873	329.6	305.4	353.6
Age-adjusted rate **			334.8	312.0	358.9			569.3	534.6	605.7			304.9	282.9	328.7

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

PRELIMINARY FINDINGS

Table 5. Incidence rates of first hospitalizations for cholelithiasis (ICD-9 574.0-574.5) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

| | Air Force enlisted | | | | | Army enlisted | | | | | Navy enlisted | | | | |
| | | | | 95 Percent confidence interval # | | | | | 95 Percent confidence interval # | | | | | 95 Percent confidence interval # | |
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper
17-19	7	22496	31.1	12.5	64.1	12	26964	44.5	23.0	77.9	12	24495	49.0	21.4	76.4
20-21	64	41171	155.5	120.6	200.2	55	52044	105.7	78.4	139.5	60	42167	142.3	106.3	178.2
22-24	71	56764	125.1	98.4	158.6	95	65802	144.4	116.8	179.1	106	53648	197.6	160.0	235.1
25-29	123	67750	181.6	151.5	217.4	108	74195	145.6	119.1	177.6	83	53679	154.6	121.3	187.8
30-34	85	47644	178.4	143.4	222.1	72	51040	141.1	110.8	179.2	80	27818	287.6	224.5	350.5
35-39	54	32769	164.8	124.1	216.2	54	27890	193.6	143.7	255.6	38	11958	317.8	216.8	418.4
40-44	22	10322	213.1	133.6	321.8	22	8439	260.7	163.5	393.7	10	3278	305.1	117.0	492.0
45-61	3	1085	276.5	57.0	807.4	9	2423	371.5	170.1	705.9	6	831	722.1	150.5	1290.0
Total	429	280001	153.2	139.3	168.6	427	308797	138.3	125.3	152.7	395	217873	181.3	163.4	199.2
Age-adjusted rate **			149.4	135.4	164.9			137.0	124.1	151.2			189.9	170.7	211.2

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

PRELIMINARY FINDINGS

Table 6. Incidence rates of first hospitalizations for cholecystitis (ICD-9 575.0-575.1) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

| | Air Force enlisted | | | | | Army enlisted | | | | | Navy enlisted | | | | |
| | | | | 95 Percent confidence interval # | | | | | 95 Percent confidence interval # | | | | | 95 Percent confidence interval # | |
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper
17-19	1	22496	4.5	0.1	24.8	6	26964	22.3	8.2	48.6	5	24495	20.4	2.8	38.0
20-21	13	41171	31.6	16.8	54.0	11	52044	21.1	10.5	37.8	11	42167	26.1	10.7	41.3
22-24	14	56764	24.7	13.5	41.4	17	65802	25.8	15.0	41.3	14	53648	26.1	12.5	39.6
25-29	13	67750	19.2	10.2	32.8	33	74195	44.5	30.0	63.6	13	53679	24.2	11.1	37.3
30-34	20	47644	42.0	25.7	64.7	14	51040	27.4	15.0	46.0	5	27818	18.0	2.4	33.4
35-39	10	32769	30.5	14.7	56.2	6	27890	21.5	7.9	46.9	13	11958	108.7	49.8	167.3
40-44	5	10322	48.4	15.7	112.9	2	8439	23.7	2.9	85.6	4	3278	122.0	4.6	238.8
45-61	2	1085	184.3	22.3	665.4	1	2423	41.3	1.0	230.0	0	831	0.0	0.0	0.0
Total	78	280001	27.9	22.2	34.9	90	308797	29.2	23.6	36.2	65	217873	29.8	22.6	37.1
Age-adjusted rate **			27.2	21.4	34.5			29.0	23.5	36.0			31.1	23.9	40.4

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

PRELIMINARY FINDINGS

Table 4. Incidence rates of first hospitalizations for kidney infections (ICD-9 590.0-590.9) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

	Air Force enlisted					Army enlisted					Navy enlisted				
				95 Percent confidence interval #					95 Percent confidence interval #					95 Percent confidence interval #	
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper
17-19	132	22496	586.8	492.7	698.5	164	26964	608.2	519.4	712.2	57	24495	232.7	172.3	292.9
20-21	169	41171	410.5	352.0	478.6	286	52044	549.5	484.7	623.1	122	42167	289.3	238.0	340.6
22-24	158	56764	278.4	237.4	326.3	263	65802	399.7	352.5	453.3	130	53648	242.3	200.6	283.9
25-29	140	67750	206.6	174.4	244.7	210	74195	283.0	245.6	325.7	83	53679	154.6	121.3	187.8
30-34	60	47644	125.9	97.0	163.7	105	51040	205.7	168.3	251.0	44	27818	158.2	111.4	204.7
35-39	30	32769	91.6	61.8	130.9	43	27890	154.2	110.1	209.7	17	11958	142.2	74.7	209.2
40-44	11	10322	106.6	53.2	190.8	10	8439	118.5	56.9	218.0	6	3278	183.0	38.1	327.0
45-61	2	1085	184.3	22.3	665.4	4	2423	165.1	44.9	422.7	1	831	120.4	0.0	240.7
Total	702	280001	250.7	232.7	270.2	1085	308797	351.4	330.0	373.9	460	217873	211.1	191.8	230.4
Age-adjusted rate **			262.8	243.9	283.3			354.0	332.4	376.7			197.3	179.7	216.6

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

PRELIMINARY FINDINGS

Table 8. Incidence rates of first hospitalizations for pelvic inflammatory disease (ICD-9 614.0-614.9) in active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

	Air Force enlisted					Army enlisted					Navy enlisted				
				95 Percent confidence interval #					95 Percent confidence interval #					95 Percent confidence interval #	
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper	No. of cases †	No. of person-years at risk	Rate	Lower	Upper
17-19	123	22496	546.8	456.4	654.8	316	26964	1171.9	1045.3	1313.7	97	24495	396.0	317.2	474.6
20-21	200	41171	485.8	421.7	559.1	476	52044	914.6	833.2	1004.2	219	42167	519.4	450.6	588.1
22-24	258	56764	454.5	401.6	514.5	598	65802	908.8	831.6	993.3	289	53648	538.7	476.6	600.7
25-29	497	67750	733.6	671.1	802.0	922	74195	1242.7	1163.2	1327.2	423	53679	788.0	712.9	863.1
30-34	395	47644	829.1	750.6	916.0	716	51040	1402.8	1301.8	1512.2	317	27818	1139.6	1014.0	1264.9
35-39	262	32769	799.5	707.1	904.2	390	27890	1398.4	1257.2	1555.0	172	11958	1438.4	1223.3	1653.0
40-44	67	10322	649.1	506.6	830.2	101	8439	1196.9	979.1	1460.2	57	3278	1738.8	1287.3	2188.8
45-61	7	1085	645.2	258.7	1329.0	19	2423	784.2	472.1	1223.4	4	831	481.4	18.2	942.1
Total	1809	280001	646.1	616.4	675.8	3538	308797	1145.7	1075.8	1219.0	1578	217873	724.3	688.5	760.0
Age-adjusted rate **			630.1	591.7	670.4			1141.3	1071.7	1214.3			752.9	707.0	801.1

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).
‡ Approximately three percent of diagnoses are unconfirmed.
§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.
# Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.
** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

PRELIMINARY FINDINGS

Table 7. Incidence rates of first hospitalizations for endometriosis (ICD-9 617.0-617.9) In active-duty enlisted Air Force, Army and Navy women per 100,000 person-years, 1990-1994

	Air Force enlisted					Army enlisted					Navy enlisted				
Age in years	No. of cases †‡	No. of person-years at risk §	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper	No. of cases †	No. of person-years at risk	Rate	95 Percent confidence interval # Lower	Upper
17-19	28	22496	124.5	82.8	180.5	43	26964	159.5	113.9	216.9	25	24495	102.1	62.1	141.8
20-21	70	41171	170.0	133.5	215.9	123	52044	236.3	196.8	283.6	72	42167	170.8	131.3	210.1
22-24	145	56764	255.4	216.2	301.6	182	65802	276.6	238.4	320.9	102	53648	190.1	153.2	227.0
25-29	247	67750	364.6	321.3	413.8	319	74195	430.0	383.6	482.0	191	53679	355.8	305.3	406.2
30-34	214	47644	449.2	391.6	514.9	262	51040	513.3	452.7	582.1	187	27818	672.2	575.8	768.4
35-39	183	32769	558.5	481.9	647.1	184	27890	659.8	568.7	765.4	150	11958	1254.4	1053.5	1454.8
40-44	56	10322	542.5	411.7	709.6	62	8439	734.7	565.7	955.1	46	3278	1403.3	997.7	1807.3
45-61	9	1085	829.5	379.9	1576.0	15	2423	619.1	346.7	1021.5	6	831	722.1	150.5	1290.0
Total	952	280001	340.0	318.8	362.4	1190	308797	385.4	361.9	410.1	779	217873	357.6	332.4	382.6
Age-adjusted rate **			319.9	299.4	341.7			381.4	358.1	405.8			395.2	366.7	426.0

† Services are reporting cases of their own personnel admitted to their own hospitals (MTF's).

‡ Approximately three percent of diagnoses are unconfirmed.

§ Air Force approximated person-years at risk by summing end-of-year annual strength for the five years, 1990-1994.

Service-specific variations of the Haenszel et al approach to calculating confidence intervals for rarely occurring events using the Poisson distribution.

** Age adjustment was performed using the indirect standardization method. For this study, the active-duty enlisted female military population was used as a standardized population.

Acknowledgment

References

1. U.S. Congress. 103rd Congress, 1st Session. H.R. 2401. National Defense Authorization Defense Authorization Act. Subtitle D - Women's Health Research. Washington, D.C.: U.S. Government Printing Office, 1993.

2. Institute of Medicine: Recommendations for Research on the Health of Military Women, Washington, D.C., National Academy Press, 1995.

3. Hoiberg A, White JF: Health status of women in the armed forces. Naval Health Research Center, 1989, NHRC Report no. 89-37.

4. Nice DS, Hilton S: Sex differences and occupational influences on health care utilization aboard U.S. Navy ships. Naval Health Research Center, 1993, NHRC Report no. 93-13.

5. Last JM, Wallace RB, eds: Maxcy-Rosenau-Last-Public Health and Preventive Medicine, 13th ed, Norwalk, Appleton & Lange, 1992.

6. Isselbacker KJ, Braunwald E, Wilson JO et al, eds: Harrison's Principles of Internal Medicine, 13th ed, New York, McGraw-Hill, Inc., 1994.

7. Strickland BR: Sex-related differences in health and illness. Psychology of Women Quarterly 1988; 12:381-399.

8. Fogel CI: Fertility control. In CI Fogel & NF Woods, eds: Health care of women: a nursing prospective, pp 484-523, St. Louis, Mosby, 1981.

9. Garland FC, Gorham ED, Garland CF, et al: Testicular cancer in US Navy personnel. Am J Epidemiol 1988; 127:411-414.

10. Shackford SR, Mackersie RC, Holbrook TL, et al: The epidemiology of tragic death: a population-based analysis. Arch Surg 1993; 128:571-575.

11. Burt CW, ed: Injury-related visits to hospital emergency departments: United States, 1992. Advance Data 1995; 261:1-15.

12. U.S Department of Health and Human Services: The International Classification of Diseases, 9th Revision, Clinical Modification: ICD-9-CM.

13. Lilienfeld AM, Lilienfeld DE: Foundations of Epidemiology, 2nd ed., New York, Oxford University Press, 1980.

14. Garland FC, Helmkamp JC, Gunderson EKE, et al: A guide to the computerized medical data resources of the Naval Health Research Center. Naval Health Research Center, 1987, NHRC Report no. 87-13.

15. Last J, ed: A Dictionary of Epidemiology, 2nd ed., New York, Oxford University Press, 1988.

16. Mausner JS, Kramer S: Epidemiology - An Introductory Text, 2nd ed., Philadelphia, W.B. Saunders Company, 1985.

17. Fleiss JL: Statistical Methods for Rates and Proportions, 2nd ed., New York, John Wiley & Sons, 1981.

18. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, 4th ed., Washington, DC, American Psychiatric Association, 1994.

19. Gomberg ES, Franks V (eds): Gender and Disordered Behavior: Sex Differences in Psychopathology, New York, Brynner/Mazel Publishers, 1979.

20. Briscoe M: Sex differences in psychological well-being. Psychological Medicine, Monograph Supplement 1982; Suppl 1:1-46

21. Nolen-Hoeksema S: Sex differences in unipolar depression: evidence and theory. Psych Bulletin 1987; 101:259-282.

22. Harris T, Surtees P, Bancroft J: Is sex necessarily a risk factor to depression? Br J Psychiatry 1991; 158:708-712.

23. Golomb M, Fava M, Abraham M, et al: Gender differences in personality disorders. Am J Psychiatry 1995; 152:579-582.

24. Hooten TM, Stamm WE: Management of acute uncomplicated urinary tract infection in adults. Medical Clinics of North America 1991; 75:339-357.

25. Remis RS, Gurwith MJ, Gurwith D et al: Risk factors for urinary tract infections. Am J Epidemiol 1987; 126:685.

26. Kunin CM, McCormack RC: An epidemiologic study of bacteria and blood pressure among nuns and working women. N Engl J Med 1968; 278:635.

27. Leiner S: Recurrent urinary tract infections in otherwise healthy adult women. Rational strategies for workup and management. Nurse Practitioner 1995; 20:48, 51, 52, 54-56.

28. Hedman P, Ringertz O: Urinary tract infections caused by staphylococcus saprophyticus. A matched case control study. J Infection 1991; 23:145-153.

29. Tullus K, Jacobson SH, Katouli M, et al: Relative importance of eight virulence characteristics of pyelonephritogenic Escherichia coli strains assessed by multivariate analysis. J Urology 1991; 146:1153-1155.

30. Andriole VT: Urinary Tract infections in the 90s. Pathogenesis and management. Infection, 1992; 20(Suppl 4):S251-S256.

31. Dalet F, Segovia T, Del Rio G: Frequency and distribution of uropathogenic E. coli adhesins: a clinical correlation in over 2000 cases. European Urology 1991; 19:295-303.

32. Brauner A, Jacobson SH, Kuhn I: Urinary Escherichia coli causing recurrent infections--a prospective follow-up of biochemical phenotypes. Clin Nephrology 1992; 38:318-323.

33. Garland FC, Shaw EK, Gorham ED, et al: Leukemia in US Navy enlisted personnel. Naval Health Research Center, 1989, NHRC report no. 89-2.

34. Palinkas LA, Pineda TS, Burr RG, et al: Ten-year profile of infectious and parasitic disease hospitalizations in the U.S. Navy. Milit Med 1990; 155:401-406.

35. White MR, McNally MS: Morbidity and mortality in U.S. Navy personnel from exposures to hazardous materials, 1974-85. Milit Med 1991; 156:70-73.

36. Gorham ED, Garland FC, Barrett-Connor E, et al: Incidence of insulin-dependent diabetes mellitus in young adults: experience of 1,587,630 US Navy enlisted personnel. Am J Epidemiol 1992; 138:984-987.

37. Garland FC, Garland CF, Doyle EJ, et al: Carpal tunnel syndrome and other cumulative trauma disorders of the arm and hand and occupation in U.S. Navy enlisted personnel. Naval Health Research Center, 1993, NHRC report no. 93-7.

38. Stout RW, Mitchell SB, Parkinson MD, et al: Viral hepatitis in the U.S. Air Force, 1989: an epidemiological and serological study. Aviation, Space, and Environmental Medicine 1994; 65:A66-70.

Appendix E - Data Mapping Tables

Air Force Variables Mapped to Navy Variables

The following variables were cross-mapped from Navy and Air Force databases, and were used in deriving and sorting inpatient data for the Epidemiological Study.

Air Force Variable	Description	Navy Variable	Description	Comments
DT-ADM-YR	date of admission- year	INP_ADMIT	the date on which the patient was first admitted as an inpatient for this hospitalization	Navy format CCYYMMDD
DT-ADM-MO	date of admission- month	"		
DT-ADM-DY	date of admission - day	"		
SSN	social security number	SSN	social security number	used only to identify individuals to isolate initial hospitalization episodes for each diagnosis
BENE-CAT	beneficiary category for hospitalized individual	INP_PCAT	indicates the branch/status of the patient	see codes; branch of service part of code
ADMN-AGE	age at date of admission	INP_AGE	age at date of admission	
SEX	gender of individual	DEM_SEX	gender of individual	see codes
PAY-GRADE	pay grade	INP_PAY	pay grade	see codes
DT-INT- ADMN-YR	year of initial admission for current hospitalization episode			initial admission date for current episode of care (allows for transfers from other MTFs)
DT-INT- ADMN-MO	month of initial admission for current hospitalization episode			
DT-INT- ADMN-DY	day of initial admission for current hospitalization episode			
DT-DISP-YR	year of disposition for current hospitalization episode	INP_DISCG	date of discharge from the hospital	

DT-DISP-MO	month of disposition for current hospitalization episode	"		
DT-DISP-DY	day of disposition for current hospitalization episode	"		
DIAG-NBR	indicates the order number of the discharge diagnosis for the same hospitalization episode			the Navy event code (for event codes 601-610) indicates the order number of the diagnosis for the same hospitalization
DIAG-CODE	ICD-9-CM diagnostic codes	INP_DIAG	NHRC diagnostic codes	

Code Mapping Of Inpatient Data

Air Force Beneficiary codes must be combined with pay grade codes to map to Navy Patient Category codes

Air Force Beneficiary Code + Pay Grade Code	Code Description	Navy Patient Category Code	Code Description
N11 (+ E1-E9 codes 31-39)	Navy AD enlisted	1	Navy enlisted
N11 (+ officer codes 01-10)	Navy AD commissioned officer	2	Navy officer
N11 (+ W1-W4 codes 21-24)	Navy AD Warrant Officer	3	Navy Warrant officer
N11 + CD	Navy AD	4	Aviation Cadet
N14	Naval Academy cadet/midshipman	5	Academy cadet/midshipman
M11 (+E1-E9 codes 31-39)	Marine AD enlisted	6	Marine Corps enlisted
M11 (+officer codes 01-10)	Marine AD commissioned officer	7	Marine Corps officer
M11 (+ W1-W4 codes 21-24)	Marine AD Warrant Officer	8	Marine Corps warrant officer
M11 + CD	Marine AD cadet	9	Marine Corps aviation cadet
F11 (+E1-E9 codes 31-39)	Air Force AD enlisted		
F11 (+officer codes 01-10)	Air Force AD officer		
F14	Air Force Academy cadet		

Sex codes

Air Force code	Description	Navy code	Description
		0 or missing	not reported
M	male	M	male
F	female	F	female

Appendix F - DMED Standard Data Set and Technical Solutions Forum

The DMED Data Standardization Workshop

During December 1995, the Air Force project team sponsored a data standardization workshop and technical solutions forum to explore ways to implement a common data infrastructure to support epidemiological research. This conference was a continuation of a dialogue among the services to create a Defense Medical and Epidemiological Database, a concept which evolved from the feasibility study performed by OPHSA and NHRC during the course of this DWHRP project. The basic premise of the DMED is that each service will maintain its own geographically separate database, but will make available via remote access common "sharable" data that have been standardized across the services. This "DMED data view" of each database will form the foundation of the Tri-service common data infrastructure. The DMED Data Set is a starting point for the three services to identify and standardize data to be shared among them. This data set will provide semantically-equivalent data from all three services, which will support a variety of studies that otherwise could not be done without a great deal of time and expense.

The workgroup established requirements for the initial DMED data set as following:

the data set would be limited to the "low hanging fruit" *i.e.*, data that are accessible to all three services and available for relatively easy inclusion at the earliest opportunity;
the data set would focus on personnel, demographic, and medical event subsets, but would be inclusive enough to answer a variety of epidemiological questions, especially those pertaining to women's health issues;
the data set would initially cover six years of data from 1989 to 1995;
the data set would include only active duty members (no data on reserves unless on active duty during the period covered).

Each service will be responsible for their own data maintenance. Data in the DMED will meet quality standards approved by the DWHRP Advisory Committee on Research Databases. Conventions and definitions used with DMED data will also be standardized.

Standardized Data Sources

Some variables have already been standardized across the services, but some variables are service-specific requirements and cannot be standardized. The Defense Manpower Data Center (DMDC) provides demographic, personnel, and career data on all active duty members. The Inpatient Data System (IPDS) for each service uses the Standard Inpatient Data Record (SIDR) reporting format for all members hospitalized at a military medical treatment facility (MTF), and also includes data on active duty members hospitalized at civilian hospitals.

The advantage of using standardized sources is that it will facilitate the standardization process across the services. The DMDC uses DoD Occupation Codes, so that cross-service comparisons can be made of occupations. DMDC also offers the Inter-service Separation Code. The SIDR provides codes that standardize the type of admission, patient category, and so on.

While reporting formats across services have been standardized, these data elements may not be DoD Standard Data Elements. Service-specific code tables are still included in many standardized formats, and, for many variables, there are no DoD Standard Data Elements.

Some service-specific tables may have data value (*e.g.,* the Air Force Duty Status and the Navy Accounting Category Code) and may be comparable, but they have not been standardized or cross-referenced. Data elements should be mapped to a common standard (i.e., DoD Standard Data Element) in order to have semantic equivalency.

Non-Standard Data

Functional experts should determine if data can be mapped to a common standard. The functional work group can request a modification of an existing standard, or propose the creation of a new standard, if no DoD Standard Data Element exists that will meet the need.

Accelerating the Standardization Process

The Rapid Data Standardization Guide helps streamline the process, and access to Worldwide Web resources can allow the services to collaborate more readily and reduce travel needs. People with expertise in this area are available to the working group.

Definition of the DMED Data Set

Prior to the conference, the Air Force team had examined some of the sources of data available for the database definition effort, including SIDR (Standard Inpatient Data Record) data and DMDC personnel and demographic data. These data were compared with data currently available to the Army and Navy, and a "strawman" data set was constructed for the workgroup to evaluate. The DMED "strawman" data set included, and was initially limited to, subsets of data that were currently available from the Air Force prototype research database, Navy longitudinal database (but currently no Navy officer and Marine Corps data), and the Army Medical Surveillance system. Some of these data had been cross-mapped, but not formally validated. Discussion also focused on other potentially comparable data in the service databases, which were not standardized, but which may be useful for the DMED.

A preliminary list of Air Force user requirements by functional area was distributed to the group. The list was compiled from interviews with researchers at OPHSA, the Air Force Surgeon General's office (data for the "Skunkworks" project), and from input generated by user focus groups. The DMED functional workgroup examined these requirements in terms of what data are available and what cases can be created to demonstrate DMED capabilities and uniqueness. As user representatives, the participants evaluated the current availability of the proposed data, the potential usefulness of the data for and set priorities

It was discussed that the list of user requirements might reflect only data variables that are available, and others might be equally valid, but unavailable to researchers. These would not have been included on the list. The group determined that FY 1996 goals should be kept in mind when evaluating which fields to include in the DMED.

Data for the DMED data set were organized into three categories:
- "Person" data, which does not change, including SSN, name, date of birth, etc.
- Demographic data, including duty location, unit identifier code, etc., which does change
- Medical events (SIDR) data, including hospitalizations, treatment outcomes, etc.

Considerable discussion ensued about which data elements should be included in which category, including the following comments:
- events can include personnel events (a list of events in the Navy database)
- the denominator needs definition (for example, dependents are not included; reserves serving on active duty may be included)
- personnel actions, location, and rank can be included in the demographic category

- marital status is an event and should be shown as such (i.e., tracking changes of marital status)
- for the Army and Navy, changes create a new record in the longitudinally tracked elements; if there are no changes, only the date code is updated
- tables can be kept for denominator counts

The group considered the following constraints and criteria for considering the DMED data set. Data selected should be:
- Unique (showcase the system's capabilities and support demonstration studies)
- Valuable (be valued and validated by researchers, sponsors, operational arena)
- Low-hanging fruit (extant and available to all services by January 1996)
- Within the time scope of five years (back to January 1989)
- Limited to active duty members, including reserves serving on active duty (differentiate between regular and reserve)

It was emphasized that data deemed valuable but not initially available would be built into future development phases of the DMED. Explicit plans to gather such data in the future should elicit the funding to do so.

There was a discussion of DMDC as a common data source for demographic and personnel data, ensuring consistency between services. The Navy Database Administrator indicated that the Navy takes tapes with full records, and receives data from 13 sources. He stated he would have to compare his current data with DMDC data to determine whether to go with DMDC, and agreed to perform this analysis for one calendar quarter. The Army representatives pointed out that having DMDC as a common data source would standardize and make data consistent. The Air Force team suggested it would be up to each service whether to use DMDC data alone, use DMDC data with footnotes, or go to another data source. DMDC uses common reporting formats for the different services, but there may be service-specific data in some of the fields. Much of these data have not been cross-mapped across the services. One example is the occupational specialty code (NEC, MOS, AFSC). Although the codes themselves have not been cross-mapped, DMDC created an occupational conversion index, which groups similar specialties across the services. The DoD Occupational Group variable should be included, to allow for cross-service comparisons of the members' occupations.

DMED Data Elements (FY 1996)

The group as a whole then went through the list of user requirements, and determined, through consensus, which were to be included in the DMED. Following is a list of those elements agreed upon by the group as variables that were pertinent for the DMED in FY 1996. They are presented by category:

"Person" Data
FMP*
SSN*
Personal ID*
Name*
Date of Birth (YYYYMMDD)
Sex
Race/ethnicity
BASD (length of service; adjusted)?
DoD Loss Code
Education level
Duty status?
Inter-service separation code?

Demographic Data
Grade
Marital status*
Duty MOS/AFSC/NEC
DoD Occupation Group
Unit Identification Code
Unit ZIP Code
MACOM/MAJCOM
Component
Primary MOS/AFSC/NEC?

Medical Events (SIDR) Data
Date of event (admission/discharge dates)
Medical Treatment Facility (MTF) location
Diagnosis Code (8 codes maximum)
Procedure Code (8 codes maximum)
Disposition
Cause of Injury
Sick days
Autopsy (yes/no indicator)
Register Number
* = Items that need further discussion for clarification.
? = Items that were questionable.

Items under the Demographic Data and Medical Events (SIDR) Data are longitudinal and are updated throughout the member's career. SSN, FMP, and name data will not be directly accessible through the DMED. However, changes to name and SSN will be captured as events. FMP is used to identify the active duty member, or "sponsor" for the purpose of matching hospitalization events from the SIDR database. The personal ID will link those three to identify an individual for research purposes. Dr. Templeton noted that race/ethnicity is a key variable within DoD.

Future DMED Data Elements (Beyond FY 1996)

In light of the previous discussion on expanding the DMED to include other elements, a list was compiled of items that might be gathered in the future:

- Health risk assessment data
- Reportable diseases data
- Outpatient treatment information
- Physical and medical evaluation board information
- Residual disabilities data
- TDY assignments
- Immunizations
- Lost duty days (other than hospital)
- Non-military MTF care information
- Dependent information (including DOB)
- Fitness data
- Death data
- Health Enrollment Assessment Review (HEAR) data (Tricare)
- Veterans Affairs data
- Major deployment information (dates and locations)

It was generally agreed that some of these data would take considerable time to acquire, while others should be obtainable within a reasonable time and sources should be pursued in the ensuing phases of the DMED. No time frame for acquisition was determined for any individual item, although information on major deployments was identified as a priority item for the future.

The list thus generated from the user requirements was compared then to the "strawman" data set.

In a discussion of medical-event items, it was suggested that if the Navy used SIDR data, it would allow compatibility across the three services and simplify the process of creating the DMED. The Navy is currently using CHCS data received from NMIMC. Although there are similarities to SIDR, some of the lookup tables in the Navy's inpatient event are markedly different. The Navy DBA agreed to look into this.

The group then focused on those items noted with question marks on the initial list of data elements:

BASD	It was determined that this is an adjusted date of service; retained
Duty status	The question centered on what this code reveals; retained
Inter-service Separation Code	It was determined that this code deals with gains and losses between services, although it does list specific reasons for the separation in question; consideration was given to using only the DoD Loss Code, but further research is required; retained temporarily
Primary MOS/AFSC/NEC	It was determined that this element needs study, but that the DoD Occupational Group code would be a common denominator.

***

PULHES	It was determined that this data is only acquired at accession, and the element was dropped from the list

A discussion was held on ensuring the equivalence of DMDC data. The group felt that efforts should be made to ensure that the service data going to DMDC fit the definitions described for the defined data elements. A comfort level could be established with DMDC data if there is semantic equivalency. DMDC should be reliable and that great differences would point out any errors. Each service could evaluate the results and make corrections as appropriate. Official data from other sources could also be used to check the reliability and accuracy of the data, such as figures on force strength.

After discussing the issue, it was decided by consensus that all three services would acquire monthly data tapes from DMDC going back six years to January 1989 from which to build the DMED, and that the effort should be completed by the end of the fiscal year. It was undecided whether one service would request tapes for all three, or if each service would request its own. A decision on this matter will be based in part on whether DMDC produces a single tape containing data for all three services, or individual tapes for each. Dr. Templeton suggested obtaining the data through CEIS, as appropriate.

Technical Solutions Forum

A number of topics were discussed during this portion of the conference. Representatives from Sybase and Oracle had been invited to do presentations of their database systems and middleware products which could be used in the implementation of the DMED communications infrastructure.

The group discussed the issues of interactive or unplanned inquiries versus planned queries, and distribution vs. centralization. The Army representative indicated each node would be able to perform the same interactive, sophisticated query if the schema were the same, but DMED need not provide that capability initially. The group determined that eventually, cross-service access will be necessary for data mining.

Data assembly, transfer, and distribution will occur at each node, and semantic equivalency will exist across DMED data from all three services. The primary issue is scalability. In order to demonstrate the value of DMED, access will be limited and controlled the first year; there will be a need to scale up in later years. Given that scenario, which approach, centralization or distribution, seems the most desirable?

Sybase representatives commented that structuring the database tables and indices for planned queries can produce very short response times, while interactive queries cannot be part of the planning process, thus response times lengthen and productivity can drop. He then covered the options available for handling workloads with Sybase products (SQL Anywhere, System 10, System 11, Sybase MPP, and Sybase IQ). He noted that Sybase IQ, which uses a different architecture, is engineered for data warehousing, and is tuned for interactive data warehousing.

Omni Connect could reside on all three nodes, or anywhere on a network.

The DMED development group should look ahead to the need for complex queries, and the need to bring back a subset of data to analyze locally.

Simple queries can require complex linkages, and that, while in the short term the approach ought to be simple, scalability should be maintained.

Multiple data locations introduced a more rigid programmatic approach, while a single location offered more flexibility, and that the data could move.

Sybase IQ could be added at any or all nodes, and would compress the data into "bitmaps." (Bitwise indices are created, which facilitate rapid response to queries. This product, however, is stand alone and separate from the original Sybase database.) This would reduce processing impact, speed access, and allow movement of data in pieces. He also stated that a centralized database would be preferable in the long term, if interactive data warehousing is desirable. Baylor College of Medicine's statewide immunization study is an example of a client using Sybase IQ for a purpose similar to DMED's.

The restraints on using Sybase IQ are based on traffic on the network and limitations on users. The product could handle databases from one to 100 gigabytes, but that it might be impractical for those less than 12 gigabytes.

Information was presented on meta data covering four areas: on-line analysis processing (OLAP), client-side, middleware, and server-side. Sybase also supports multi-dimensional database analysis (along four dimensions) through partnerships with various vendors. OLAP is most often used with financial and marketing applications and provides trend analysis support, which could be useful for longitudinal studies of treatment interventions.

Multi-dimensional analysis could be used with existing software, would benefit more sophisticated DMED users, and often uses a nested spreadsheet as the user interface.

The Air Force agreed to develop the proposed DMED database into a standard, logical design to present to the other services.